



(43) International Publication Date 30 May 2002 (30.05.2002)

PC1

(10) International Publication Number WO 02/42468 A2

(51) International Patent Classification⁷: C12N 15/52

(21) International Application Number: PCT/US01/44306

(22) International Filing Date:

26 November 2001 (26.11.2001)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

60/253,395

27 November 2000 (27.11.2000) US

(71) Applicant (for all designated States except US): GERON CORPORATION [US/US]; 230 Constitution Avenue, Mcnlo Park, CA 94025 (US).

(72) Inventor; and

(75) Inventor/Applicant (for US only): SCHIFF, Michael, J. [CA/US]; 808 Coleman Avenue, Apt. 19, Menlo Park, CA 94025 (US).

(74) Agent: EARP, David, J.; Geron Corporation, 230 Constitution Drive, Menlo Park, CA 94025 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

 without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: GLYCOSYLTRANSFERASE VECTORS FOR TREATING CANCER

| Marmoset ul, 3GT human pseudogene | MAVXCKVILSMLVVSTVIVVFWEYINSPEGSFLWIYHSKNPEV-DDSSAQXDWWFPGWFWWCIHNYQQEE |
|--|--|
| sheep | |
| Bovine | |
| Pig | R.VLMLFDGRGST.SH |
| Mouse | L.IVV.RI |
| | |
| Consensus al. 3GT | |
| Humanized al. 3GT | |
| • | |
| hu B transferase | . AEVLRTLACKPKCHALRPMILFL.MLVLVL. CYCVLSPRSLMPCSL |
| hu A transferase | . AEVERTLAGKPKOHALRPMILFE. MLVLYL. GYGVESPRSLMPGSL |
| | |
| Marmoset 01,3GT human pseudogene | EDTDX-EKGREEEQKKEDDTTELRLWOWFNPKKRPEVMTVTQWKAPVVWEGTYNKAILENYYAKQKITVG |
| sheep | DEDV0EEQRK.DSK.K.SFV.M.D |
| Bovine | DGDINERNE-SK.K.SFY.M.K |
| Pig | AIGNRNRGP.VEV.I.RR.V.D |
| Mouse | NVEGRRGRNG.RIE.PQND.LPIDT.LKTL |
| | |
| Consensus al, 3CT Humanized al, 3CT | V.M.PIR.V.D |
| HUMANIZED ILL. SCI | |
| bu B transferase | .RCFCHAVREPDHLORVSLPRMVYPOPKVLTPC.KD.LVP.LIF.LDNEDFRL.NT.I. |
| hu A transferase | .RGFCMAVREPDHLORVSL PRMVYPOPKVLTPC.KD.LVP.LIF.IDNEOFRL.NT.I. |
| | |
| Marwoset al.3GT | LTVFAIGRYIEHYLEEFVTSANRYFMYGHKYIFYVMYDDVSKAPFIELGPLRSFKVFEVKPEKRWODISM |
| human pseudogene | . NO I |
| sheep | V |
| Bovine | v |
| Pig | VLITIIRA.LI.S |
| Mouse | V.KO.LEO4RIT.RM.VVH.NH.LQIRS |
| Consensus al.3CT | V.KRIRM.LI.SV. |
| Humanized al.3GT | V.K |
| mmai:1200 01,361 | |
| hu B transferase | KK.VA-F.KL.LET.EKHR.HYFT.QPAAV.RVTTC.QLS.LCAYV |
| hu A transferase | KK.VA-F.KL.LET.EKHR.HYFT.OLAAV.RVTTG.OLS.LRAYV |

(57) Abstract: This disclosure provides a system for specifically killing cancer cells which can be used in the course of human therapy. Vectors of the invention comprises an encoding sequence for a glycosyltransferase, under control of a tumor or tissue specific transcriptional control element, such as the promoter for telomerase reverse transcriptase. Exemplary glycosyltransferases are the A or B transferase enzymes, which cause the cancer cells to express ABO histo blood group allotypes against which humans have naturally antibody. This provides for ongoing surveillance for newly emerging cells with a malignant phenotype.



GLYCOSYLTRANSFERASE VECTORS FOR TREATING CANCER

TECHNICAL FIELD

This invention relates generally to the field of virology and cancer therapy. This disclosure provides vectors in which an encoding region for glycosyltransferase is linked to a genetic element that controls transcription in a tumor or tissue specific fashion.

REFERENCE TO RELATED APPLICATION

10

5

This application claims priority to U.S. Patent Application 60/253,395; filed November 27, 2000, pending. Where permitted, the priority application is hereby incorporated herein by reference in its entirety.

15

BACKGROUND

Many forms of cancer are intractable to traditional courses of radiation or small molecule pharmaceuticals. Considerable interest has evolved in developing gene therapy vectors as chemotherapeutic agents.

20

A broad variety of therapeutic genes are currently under investigation in preclinical and in clinical studies (Walther et al., Mol. Biotechnol. 13:21, 1999). The candidate genes have very different origins and different mechanisms of action — which include cytokine genes, genes coding for immunostimulatory molecules/antigens, genes encoding bacterial or viral prodrug-activating enzymes (suicide genes), and tumor suppressor genes.

25

Some of the putative vectors are based on adenovirus. U.S. Patents 5,631,236 and 6,096,718 (Baylor College of Medicine) cover a method of causing regression in a solid tumor, using a vector containing an HSV thymidine kinase (tk) gene, followed by administration of a prodrug such as ganciclovir. U.S. Patent 6,096,718 (Baylor College of Medicine) relates to the use of a replication incompetent adenoviral vector, comprising an HSV tk gene under control of the α -lactalbumin promoter.

30

U.S. Patents 5,801,029 and 5,846,945 (Onyx Pharmaceuticals) relate to adenovirus in which the E1a gene has been altered so as not to bind and inactivate tumor suppressor p53 or RB. This prevents the virus from inactivating tumor suppression in normal cells, which means the virus cannot replicate. However, the virus will replicate in cells that have shut off p53 or RB expression through oncogenic transformation.

35

40

U.S. Patent 5,998,205 (GTI/Novartis) pertains to a tissue-specific replication-conditional adenovirus, comprising a transcriptional regulatory sequence (such as the α-fetoprotein promoter) operably linked to adenovirus early replication gene. U.S. Patent 5,698,443 (Calydon) provides replication-conditional adenoviruses controlled by the PSA promoter. Alemany et al. (Cancer Gene Ther. 6:21, 1999) outline complementary adenoviral vectors for oncolysis. One vector contains cis replication elements and E1a under control of a tissue-specific promoter. The supplemental vector contains all other

trans-acting adenovirus replication genes. Coinfection leads to controlled killing of hepatocarcinoma cells.

International Patent Publication WO 98/14593 (Geron) describes an adenovirus construct in which the *tk* gene is placed under control of the promoter for telomerase reverse transcriptase (TERT). This gene is expressed at high levels in cancer cells of any tissue type, and the vector renders cancer cell lines susceptible to toxic effects of ganciclovir. WO 00/46355 (Geron) describes an oncolytic virus having a genome in which a TERT promoter is linked to a genetic element essential for replication or assembly of the virus, wherein replication of the virus in a cancer cell leads to lysis of the cancer cell.

5

10

15

20

25

30

35

40

Koga et al. (Hu. Gene Ther. 11:1397, 2000) propose a telomerase-specific gene therapy using the hTERT gene promoter linked to the apoptosis gene Caspase-8 (FLICE). Gu et al. (Cancer Res. 60:5359, 2000) reported a binary adenoviral system that induced Bax expression via the hTERT promoter. They found that it elicited tumor-specific apoptosis in vitro and suppressed tumor growth in nude mice.

Other vectors are based on herpes family viruses, such as herpes simplex type 1 and 2. U.S. Patent 5,728,379 (Georgetown University) relates to replication competent HSV containing a transcriptional regulatory sequence operatively linked to an essential HSV gene. Exemplary is the IPC4 gene under control of the pro-opiomelanocortin promoter.

Other vectors are based on the retrovirus family. U.S. Patent 5,997,859 and EP 702084 B1 (Chiron) pertain to replication-defective recombinant retrovirus, carrying a vector construct capable of preventing, inhibiting, stabilizing or reversing infections, cancer, or autoimmune disease. The virus directs expression of an enzyme not normally expressed in the cells that converts a compound into a cytotoxic form. Exemplary is the HSV tk gene. WO 99/08692 proposes the use of reovirus in treating cancer, particularly ras-mediated neoplasms.

These proposed therapeutic agents are not currently approved for commercial use in the United States. There is a need to develop new constructs to improve efficacy and specificity of cancer treatment.

SUMMARY OF THE INVENTION

This invention provides a system for killing cancer cells in vitro or in vivo, using a polynucleotide encoding a glycosyltransferase under control of a tumor specific or tissue specific transcriptional control element. The glycosyltransferase typically forms a determinant on the cell surface to which some or all humans have naturally occurring antibody. In this manner, cancer cells will be culled on an ongoing basis by antibody already present in the circulation, without the need to follow the vector with an effector agent.

One embodiment of the invention is a polynucleotide as already described. Suitable glycosyltransferase enzymes include but are not limited to histo blood group A or B transferase from any upper primate (particularly human), and $\alpha(1,3)$ galactosyltransferase ($\alpha(1,3)$ Gal xenoantigen.

The transcriptional control element can be a tissue specific promoter, as exemplified below. Alternatively, the control element can be a tumor specific promoter, as exemplified below. Of particular interest is the promoter for telomerase reverse transcriptase (SEQ. ID NO:1). The polynucleotide can

take the form of a viral vector (for example, adenovirus, herpes virus, or retrovirus), naked DNA, or a lipid composition (for example, a neutral or anionic lipid envelope, or a cationic liposome or micelle) that has a DNA or RNA component.

Polynucleotides of the invention can be used to prepare a medicament for human treatment, especially for conditions associated with hyperproliferation, such as cancer and other neoplasias.

Another embodiment of the invention is a polypeptide with glycosyltransferase activity, which comprises a consensus of mammalian α 1,3GT sequences, or a humanized α 1,3GT sequence, or catalytic subfragment thereof.

Also provided is a method of killing a cancer cell, comprising combining the cancer cell with a polynucleotide as already described. The invention includes a system for testing and manufacturing the glycosyltransferase vectors of this invention. The invention can be used for treating cancer in a subject by administering to the subject a polynucleotide as already described.

Other embodiments of the invention will be apparent from the description that follows.

15 DRAWINGS

5

10

20

25

30

35

40

Figure 1 is a map of adenovirus vector designated pGRN376, in which the promoter for telomerase reverse transcriptase (TERT) controls expression of the *tk* gene (Example 1).

Figure 2 is a photographic reproduction showing the effects of replication-conditional adenovirus on normal and cancer-derived cell lines (Example 2).

Figure 3 is a sequence listing comparing the human blood group A and B transferase amino acid sequences with $\alpha(1,3)$ galactosyltransferase ($\alpha1,3$ GT) of other species. A consensus version and a humanized version of $\alpha1,3$ GT are shown as SEQ. ID NOs:12 & 13. (-) represents a sequence gap; (.) indicates a residue identical with the aligned marmoset $\alpha1,3$ GT sequence (Example 3). Other sequences shown in this figure are listed in Table 2.

Figure 4 is a sequence listing comparing the marmoset α 1,3GT encoding sequence with the human α 1,3GT pseudogene. The humanized α 1,3GT encoding sequence is shown as SEQ. ID NO:16 (Example 3). The sequences shown in this figure are listed in Table 2.

DETAILED DESCRIPTION

A long-sought objective in cancer treatment is to design a therapeutic agent that effectively kills cancer cells wherever they are located, while sparing other cells in the vicinity that do not bear the malignant phenotype.

The invention described in this disclosure solves the problem by providing a therapeutic vector that encodes an enzyme that forms a target molecule on the cell surface that can be targeted by antibody in situ. Particularly effective are so-called natural antibodies that recognize features of foreign complex carbohydrates. A number of naturally occurring anti-carbohydrate antibodies are present in the circulation of humans without deliberate immunization. It is thought that these antibodies arise from cross-reacting mucins and other carbohydrate-bearing substances that people are routinely exposed to through their diet.

In one aspect of this invention, the carbohydrate targets are produced in greater abundance on tumor cells, because expression of the enzyme that makes the target is controlled by a transcriptional control element that is tumor or tissue specific. Tumor-specific targeting relies on control elements taken from genes expressed predominantly in cells that undergo repeated proliferation, or that are relatively undifferentiated. Such vectors are effective for treating a wide variety of tumor types at the primary site or elsewhere. Tissue-specific targeting relies on control elements taken from genes expressed in particular tissue types. Such vectors are especially useful for treating metastases, or tumors in which the tissue-specific element is relatively more abundant.

5

10

15

20

25

30

35

40

Treatment is effected by administering the vector systemically or locally so that it can migrate to and transfect the tumor cells causing the disease. The vector then causes expression of the new carbohydrate structure at the cell surface. This becomes a target for antibody in the circulation (or other components of the immune system, such as cytotoxic T cells, ADCC cells, or T helper/inducer cells) — which in turn leads to a number of possible effects — complement-mediated lysis, opsonization, cytotoxic killing, cytokine and interferon secretion, and inflammatory response.

This system is believed to offer two advantages over previous approaches to gene therapy for cancer.

The first advantage is that it can provide ongoing surveillance against the emergence of new malignancies. This is available when using a tumor-specific expression vector, such as the TERT promoter described below, and when the vector is capable of replication or remains expressible by the cell. In cancer cells, the vector will cause expression of the target carbohydrate, causing them to be recognized and eliminated by antibody. In cells that are not actively malignant, the vector will remain quiescent — until such time as the cell reverts to the cancer phenotype — whereupon the target carbohydrate will be expressed de novo, and the cell becomes eliminated in its turn. Since naturally occurring antibody is persistently available, there is no need to readminister an effector drug to eradicate any newly activated cancer cells.

. The second advantage is that glycosyltransferases potentially provide a second level of specificity for malignant cells. In using tumor-specific promoters to drive gene expression, there is at least a theoretical concern that the vector may also have an effect on non-cancerous cells that upregulate the promoter transiently as part of the normal replicative process of the cell. For example, TERT is expressed transiently by some actively growing stem cells, lymphocytes, and germinal tissue.

The potential second layer of specificity provided by glycosyltransferase is related to the density of carbohydrate determinants on the surface of certain types of progenitor cells. Immune lysis of cells through glycolipid antigen depends primarily on IgG antibody. The IgG molecule must span two antigenic determinants with its two combining sites in order to activate complement — binding to only one determinant (termed monogamous bivalency) is insufficient. This means there is a minimum density of determinants that must be present in order for the antibody to activate complement.

Fetal red cells bear a low density of ABO blood group determinants, attributable to paucity of branches in the oligosaccharide. This means that ABO blood group IgG antibodies can only bind monogamously (Romans et al., J. Immunol. 124:2807, 1980). If other fetal and embryonic cells express the branching enzyme in the same limited fashion, then they may also be less susceptible to complement lysis mediated by antibodies directed against any part of the same complex carbohydrate.

This theoretical rationale is provided to enhance the reader's appreciation of the invention. Those skilled in the art will appreciate that there are other advantages in the invention beyond those indicated above. This explanation is not meant to limit the claimed invention in any way.

Further explanation of the making and use of the vector constructs of the invention is provided in the sections that follow.

Definitions

5

10

15

20

25

30

35

40

The term "polynucleotide" refers to a polymeric form of nucleotides of any length. Included are genes and gene fragments, mRNA, tRNA, rRNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA and RNA, nucleic acid probes, and primers. As used in this disclosure, the term polynucleotides refer interchangeably to double- and single-stranded molecules. Unless otherwise specified or required, any embodiment of the invention that is a polynucleotide encompasses both a double-stranded form, and each of the two complementary single-stranded forms known or predicted to make up the double-stranded form.

A cell is said to be "genetically altered", "transfected", or "genetically transformed" when a polynucleotide has been transferred into the cell by any sultable means of artificial manipulation, or where the cell is a progeny of the originally altered cell that has inherited the polynucleotide. The polynucleotide will often comprise a transcribable sequence encoding a protein of interest, which enables the cell to express the protein at an elevated level. The genetic alteration is said to be "inheritable" if progeny of the altered cell have the same alteration.

A "control element" or "control sequence" is a nucleotide sequence that contributes to the functional regulation of a polynucleotide, such as replication, duplication, transcription, splicing, translation, or degradation of the polynucleotide. Transcriptional control elements include promoters, enhancers, and repressors.

Particular gene sequences referred to as promoters, like the "TERT promoter", or the "PSA promoter", are polynucleotide sequences derived from the gene referred to that promote transcription of an operatively linked gene expression product. It is recognized that various portions of the upstream and intron untranslated gene sequence may in some instances contribute to promoter activity, and that all or any subset of these portions may be present in the genetically engineered construct referred to. The promoter may be based on the gene sequence of any species having the gene, unless explicitly restricted, and may incorporate any additions, substitutions or deletions desirable, as long as the ability to promote transcription in the target tissue. Genetic constructs designed for treatment of humans may comprise a segment that at least 90% identical to a promoter sequence of a human gene. A particular sequence can be tested for activity and specificity, for example, by operatively linking to a reporter gene (Example 1).

Genetic elements are said to be "operatively linked" if they are in a structural relationship permitting them to operate in a manner according to their expected function. For instance, if a promoter helps initiate transcription of the coding sequence, the coding sequence can be referred to as operatively linked to (or under control of) the promoter. There may be intervening sequence between the promoter and coding region so long as this functional relationship is maintained.

In the context of encoding sequences, promoters, and other gene elements, the term "heterologous" indicates that the element is derived from a genotypically distinct entity from that of the rest of the entity to which it is being compared. For example, a promoter or gene introduced by genetic engineering techniques into a context; in which it does not occur in nature is said to be a heterologous polynucleotide. An "endogenous" genetic element is an element that is in the same place in the chromosome where it occurs in nature, although other gene elements may be artificially introduced into a neighboring position.

The terms "polypeptide", "peptide" and "protein" are used interchangeably to refer to polymers of amino acids of any length. The polymer may comprise modified amino acids, it may be linear or branched, and it may be interrupted by non-amino acids.

The term "antibody" as used in this disclosure refers to both polyclonal and monoclonal antibody. The ambit of the term deliberately encompasses not only intact immunoglobulin molecules, but also such fragments and genetically engineered derivatives of immunoglobulin molecules, T cell receptors, and their equivalents as may be prepared by techniques known in the art, and which retain binding specificity of the antigen combining site.

General Techniques

5

10

15

20

25

30

35

40

Methods in molecular genetics and genetic engineering are described generally in the current editions of *Molecular Cloning: A Laboratory Manual*, (Sambrook et al.); *Oligonucleotide Synthesis* (M.J. Gait, ed.,); *Animal Cell Culture* (R.I. Freshney, ed.); *Gene Transfer Vectors for Mammalian Cells* (Miller & Calos, eds.); *Current Protocols in Molecular Biology* and *Short Protocols in Molecular Biology, 3rd Edition* (F.M. Ausubel et al., eds.); and *Recombinant DNA Methodology* (R. Wu ed., Academic Press). Reagents, cloning vectors, and kits for genetic manipulation referred to in this disclosure are available from commercial vendors such as BioRad, Stratagene, Invitrogen, and ClonTech.

For a description of the molecular biology of cancer, the reader is referred to *Principles of Molecular Oncology* (M.H. Bronchud et al. eds., Humana Press, 2000); *The Biological Basis of Cancer* (R.G. McKinnel et al. eds., Cambridge University Press, 1998); and *Molecular Genetics of Cancer* (J.K. Cowell ed., Bios Scientific Publishers, 1999).

General techniques for the development, testing, and administration of biomolecular chemotherapeutics are provided in *Gene Therapy of Cancer*, Adv. Exp. Med. Biol. vol. 451 (P. Walden ed., Plenum Publishing Corp., 1998); *Cancer Gene Therapy*, Adv. Exp. Med. Biol. vol. 465(N. A. Habib ed., Kluwer Academic Pub, 2000); and *Gene Therapy of Cancer: Methods and Protocols*, Meth. Mol. Med. vol. 35 (W. Walther & U. Stein eds., Humana Press, 2000).

Effector Genes for Tumor Cell Depletion

The vectors of this invention comprise an encoding region that forms a carbohydrate determinant on the cell surface as a target for cancer cell lysis.

Exemplary are glycosyltransferases that synthesize an alloantigen or xenoantigen widely expressed on different tissue types.

In humans, an $\alpha(1,2)$ fucosyltransferase uses N-acetyl lactosamine acceptor groups on cell surface glycoproteins and glycolipids to form Fuc $\alpha(1,2)$ Gal $\beta(1,4)$ GlcNAc, which is blood group H

substance. This in turn serves as an acceptor for the ABO histo blood group transferases, which form terminal allodeterminants on the complex carbohydrate. Blood group A transferase adds GalNAc to form GalNAca(1,3)Gal (A substance). Blood group B transferase adds Gal instead to form Gala(1,3)Gal (B substance).

5

10

15

20

25

30

35

40

According to the blood group of an individual, one or both of these transferases are expressed in essentially all nucleated cells, resulting in expression of A and B substance on the cell surface. Red cells also abundantly present A and B substance, by virtue of synthesis before enucleation, and subsequent adsorption of glycolipids from plasma. Naturally occurring antibodies circulate in the blood that react against the ABO determinants that are not self-antigens. One advantage of using an ABO transferase as the effector sequence is that the H precursor substance will be available on the surface membrane of virtually any tumor.

The nucleotide and protein sequence of A transferase and B transferase are provided below. See also U.S. Patents 5,068,191 and 5,326,857. The two enzymes are close homologs of each other, differing by only a few amino acids. Another advantage of using an ABO transferase as the effector sequence is that the expressed protein is of human origin, and unlikely to be immunogenic by virtue of its similarity to another gene product expressed as a self antigen in the patient being treated.

Mammals other than humans, apes and Old World monkeys do not form H precursor substance, but instead convert the N-acetyl lactosamine acceptor into the $Gal\alpha(1,3)Gal$ determinant. $Gal\alpha(1,3)Gal$ epitope is expressed prominently on the surface of nucleated cells, including hepatic cells, renal cells, and vascular endothelium — and is the main target for the natural antibodies mediating xenograft rejection (reviewed by Joziasse et al., Biochim. Biophys. Acta 1455:403, 1999; Sandrin et al., Frontiers Biosci. 2:31, 1997).

The Gal α (1,3)Gal epitope is made by a specific enzyme, α (1,3)galactosyltransferase (α 1,3GT). In humans and other primates that don't express the Gal α (1,3)Gal product, the α 1,3GT locus is inactivated (Gailili et al., Proc. Natl. Acad. Sci. USA 15:7401, 1991). There are frameshift and nonsense mutations within the locus, turning it into a non-functional, processed pseudogene (Laarsen et al., J. Biol. Chem. 265:7055, 1990; Joziasse et al., J. Biol. Chem. 266:6991, 1991).

For use in this invention, α 1,3GT of any species can be used. A number of α 1,3GT sequences are provided below. For use in human therapy, it may be beneficial to use an α 1,3GT that differs as little as possible from the human pseudogene sequence, while retaining the same specificity. The complete marmoset α 1,3GT sequence is provided below, and can be humanized by substituting residues from the human pseudogene that do not alter the binding or catalytic site. If desired, glycosyltransferases can also be truncated down to the minimal size of the catalytically active enzyme (Henion et al., Glycobiology 4:193, 1994).

Other glycosyltransferases can also be identified for use in this invention. Candidates include transferases responsible for other carbohydrate blood group alloantigens (for example, Lewis, P, Ii blood groups). Candidates also include non-mammalian glycosyltransferases, and transferases responsible for making determinants present on embryonic cells of humans and other species that are not found on most adult cells.

The choice of a particular transferase may involve a number of considerations and routine empirical testing. One consideration is the density of determinants formed on transfected cells. As

discussed earlier, certain glycosyltransferases may synthesize a lower density of determinants on stem cells by virtue of the relative paucity of branched precursor substances on those cells. By judicious selection of the transferase, it may be possible to titrated the density of determinants formed. For example, A- and B-transferases will have exclusive access to H substance if transfected into type O cells, or will compete 1:1 with each other as counterparts. α 1,3GT is expected to produce less determinant, because it must compete in humans with the α (1,2)fucosyltransferase that forms H substance. It has been found that α 1,3GT fairs less well in this competition because of its position in the Golgi, which in turn is a function of the N-terminal membrane-anchoring domain. It is possible to switch the α (1,2)fucosyltransferase cytoplasmic domain onto α 1,3GT in order to increase the density of Gal α (1,3)Gal epitopes produced (Osman et al., J. Biol. Chem. 271:33105, 1996).

Transcriptional Control Elements for Tumor Targeting

5

10

15

20

25

30

35

40

The control element is selected with a view to the protein expression patterns in cancer cells compared with non-malignant cells that will also be exposed to the vector.

Many tumor-specific transcriptional control elements can be used in this invention. These control elements cause elevated transcription of the encoding sequence they are linked to in tumor cells of a variety of different types. Examples are promoters that control telomerase reverse transcriptase (TERT), carcinoembryonic antigen (CEA), hypoxia-responsive element (HRE), autocrine motility factor receptor (*Grp78*), L-plastin, and hexokinase II.

The promoter for TERT is exemplary. Sequence of the human TERT gene (including upstream promoter sequence) is provided below. The reader is also referred to U.K. Patent GB 2321642 B (Cech et al., Geron Corporation and U. Colorado), International Patent Publications WO 00/46355 (Morin et al., Geron Corporation), WO 99/33998 (Hagen et al., Bayer Aktiengesellschaft), and Horikawa et al. (Cancer Res., 59:826, 1999). Other TERT sequences can also be used; the mouse sequence is provided in WO 99/27113 (Morin et al., Geron Corporation). A lambda phage clone designated λ G Φ 5, containing ~13,500 bases upstream from the hTERT encoding sequence, is available from the ATCC under Accession No. 98505. Example 1 illustrates the testing and use of TERT promoter sequences in vector expression systems. Those skilled in the art will appreciate that promoter sequences not contained in λ G Φ 5 but homologous and capable of promoting preferential expression in cancer cells can be used with similar effect. For example, a TERT promoter can comprise a sequence of 25, 50, 100, or 200 consecutive nucleotides that is 80%, 90%, or 100% identical (or can hybridize under stringent conditions) to a sequence contained in SEQ. ID NO:1.

As an alternative, a transcriptional control element can be used that is tissue-specific. Constructs of this kind will cause preferential expression of the glycosyltransferase, if the level of expression of the endogenous gene is higher in tumor cells than in non-malignant tissue of the same type. They are also useful to treat tumors that have metastasized away from the primary site. Examples are promoters that control transcription of albumin (liver-specific), α-fetoprotein (AFP, liver-specific), prostate-specific antigen (PSA, prostate-specific), mitochondrial creatine kinase (MCK, muscle-specific), myelin basic protein (MBP, oligodendrocyte-specific), glial fibrillary acidic protein (GFAP, glial cell specific), and neuron-specific enolase (NSE, neuron-specific). See U.S. Patent 5,871,726 (Calydon), WO 98/39466 (Calydon), U.S. Patent 5998205 (Genetic Therapy Inc.).

<u>8</u>

Additional promoters suitable for use in this invention can be taken from other genes that are preferentially expressed in tumor cells. Such genes can be identified, for example, by differential display and comparative genomic hybridization: see U.S. Patents 5,759,776 and 5,776,683. Alternatively, microarray analysis can be performed cDNA fragments of candidate genes in a 96 or 384 well format, and then spotted directly onto glass slides. To compare mRNA preparations from cancer cells and a matched non-malignant control, one preparation is converted into Cy3-labeled cDNA, while the other is converted into Cy5-labeled cDNA. The two cDNA preparations are hybridized simultaneously to the microarray slide, and then washed to eliminate non-specific binding. Any given spot on the array will bind each of the cDNA products in proportion to abundance of the transcript in the two original mRNA preparations. The slide is then scanned at wavelengths appropriate for each of the labels, and the relative abundance of mRNA is determined. Preferably, the level of expression of the effector gene will be at least 5-fold or even 25-fold higher in the undifferentiated cells relative to the differentiated cells. Having identified transcriptional control elements of interest, specificity can be tested in a reporter transcription of a reporter gene, such as green construct where the control element is used to control fluorescence protein, secreted alkaline phosphatase, or β-galactosidase.

Formulation and Administration of Cancer Therapeutics

5

10

15

20

25

30

35

40

A number of viral vectors are suitable for cancer gene therapy according to the invention. For general principles in vector construction, the reader is referred to *Viral Vectors for Gene Therapy* (B.J. Carter, Biotechnology 1999, XVIII, 562 p. 393, 1999).

Adenovirus vectors provide transient gene expression, and can be constructed to be replication competent or replication incompetent. For general principles in adenovirus construction, see Danthinne et al., Gene Ther. 7:1707, 2000, Bilbao et al., Adv. Exp. med. Biol. 451:365, 1998, and U.S. Patents 5,631,236 (Baylor College of Medicine), 5,670,488 (Genzyme), 5,698,443 (Calydon), 5,712,136 (GenVec), 5,880,102 (Duke University), 5,994,128 (IntroGene), 6,040,174 (Transgene), 6,096,718 (Gene Targeting Corp).

Retrovirus vectors can be constructed to provide gene expression that is inheritable by progeny of the cell it infects. U.S. Patent 5,698,446 and 6,133,029 (Chiron). Vectors can also be based on viruses of the herpes family. U.S. Patent 5,728,379 (Georgetown University). Adeno-associated virus, reovirus, and a number of other viruses are also suitable.

As an alternative, the vectors of this invention can be constructed on a technology which is not virus based. Suitable are nucleic acid-lipid complexes of various kinds, where the lipid protects the nucleic acid en route to the tumor, and facilitates entry into the cell. One form is cationic liposomes or micelles. Li et al. (Gene Ther. 5:930, 1998) generally describe cationic lipid – promoter – DNA complexes for intravenous gene delivery. Another form is neutral or anionic liposomes, where the DNA is encapsulated in a lipid envelope that may express other components to inhibit non-specific uptake. U.S. Patents 5,981,501 (Inex) and 6,043,094 (Sequus/Alza). The composition may resemble an artificial viral envelope. U.S. Patent 5,766,625 (U. Florida) and WO 97/04748 (Advanced Therapies).

Also part of the invention are viral constructs in which gene expression is cell-specific, and the virus itself is replication conditional. See generally Todo et al., Cancer Gene Ther. 7:939, 2000; and WO 00/46355 (Geron). In this embodiment, the glycosyltransferase encoding region is under control of a

tissue or tumor specific control element — and a gene essential for replication or packaging of the virus is also under control of a tissue or tumor-specific control element. Genes required for replication of adenovirus include E1a, E1b, E2, and E4. Genes required for replication of HSV include ICP6 and ICP4. Glycosyltransferase expression and viral replication can be controlled by the same promoter — or they can be controlled by different promoters, providing a further level of specificity for cancer cells.

5

10

15

20

25

30

35

40

Constructs comprising different glycosyltransferase encoding regions and different regulatory control elements can be tested and compared in several different assay systems. Suitable cells for these assays include human tumor cells expressing the gene from which the regulatory control element of the virus is taken (e.g., hTERT), matched with cell lines from a similar non-malignant tissue, or a tissue expressing about the same density of acceptor substrate for the glycosyltransferase. The cells can be transduced with the test vector, with a vector not comprising the glycosyltransferase sequence (negative control), and with a vector in which the glycosyltransferase is under control of a constitutive promoter (such as CMV or PGK).

Expression of the glycosyltransferase can be detected at the RNA level by RT-PCR, and at the protein level by immunocytochemistry, according to standard techniques. Expression of the cell-surface determinant synthesized by the glycosyltransferase can be detected using epitope-specific antibody or lectin, for example, by FACS. Human type B serum contains antibodies to A substance and to the Galα(1,3)Gal xenoantigen. The "IB4" lectin from Bandeiraea (Griffonia) simplicifolia (Sigma Cat. L 3019) is specific for α-D-galactosyl residues and binds both the Galα(1,3)Gal epitope, and B blood group substance. Antigen density can be compared for vectors with different promoters and effectors in quantitative assays using labeled monovalent antibody. Monogamous bivalency (the ability or inability of specific IgG to bind by more than one combining site) can be measured in suspended cells using the antiglobulin test (Romans et al., J. Immuno). 124:2807, 1980).

Ultimately, efficacy of the constructs of this invention can be assessed by their ability to trigger complement-mediated tumor cell lysis. A panel of tumor and non-tumor lines in culture is transfected with the vector, and then exposed to a source of epitope-specific antibody plus complement. For typical vectors encoding $\alpha 1,3GT$, fresh human serum will contain sufficient antibody and complement to cause specific lysis. For typical vectors encoding an A or B transferase, fresh serum of O blood type should cause lysis. If fresh serum is not available for the product of a particular glycosyltransferase, lysis can be measured using specific antibody and guinea pig complement. Rather than measuring lysis, the cells can be treated for a brief interval and then injected into a suitable mouse model, to determine if the treatment inhibits tumor growth.

General validation of the approach and titration of virus can be confirmed using a $\alpha 1,3GT$ vector in $\alpha 1,3GT$ knockout mice. U.S. Patent 5,849,991 (Bresatch) reports mice that are homozygous for inactivated $\alpha 1,3GT$, resulting in lack of expression of $Gal\alpha(1,3)Gal$ epitope, as determined by specific antibody. A model is developed in which the mice are injected with a representative human cancer cell line, such as a glioma. After solid tumors have developed of a sizeable diameter, the mice are injected intravenously or intratumorally with the $\alpha 1,3GT$ vector. A dose of 10^5 to 10^8 pfu is the predicted test range for HSV vectors. Once the $\alpha 1,3GT$ is expressed, anti- $Gal\alpha(1,3)Gal$ in the plasma of these mice should opsonize the tumor cells, slowing tumor growth, potentially causing regression and increased survival.

Treatment of human patients having a tumor depends on the nature of the vectors available and the carbohydrate determinants naturally expressed on their cells. Patients of blood type O (~46% of the U.S. population) will have natural antibody to both A and B substance, and can be treated with a vector encoding either A or B transferase. Patients of blood type A (~38%) or B (~12%) will have natural antibody to the opposite determinant, and can be treated with a vector encoding the corresponding transferases. Patients of blood type AB (~4% of the population) will not be treatable using either vector. It is possible to use a mixture of A and B transferase vectors as a universal reagent for patients of blood types A, B, and O (~96% of the population). The lytic potential of the mixture may be somewhat reduced in blood types A and B, since the transferases will be codominantly expressed.

A universal reagent suitable for treating all ABO blood groups is a vector made using the α 1,3GT transferase. Since humans don't have the anti-Gal α (1,3)Gal epitope, essentially everyone should have naturally occurring antibody. α 1,3GT must compete in humans for the N-acetyl lactosamine acceptor substrate with the α (1,2)fucosyltransferase that makes H substance. Since α 1,3GT fairs less well in this competition because of its position in the Golgi (Osman et al., J. Biol. Chem. 271:33105, 1996), a higher density of epitope will be formed by a construct that encodes the N-terminal membrane anchoring domain of the α (1,2)fucosyltransferase fused to the extramembrane catalytic domain of α 1,3GT.

Dosage and formulation of medicaments intended for human therapy are designed based on the animal model experiments. For general guidance on formulation and testing of medicament formulations for human administration, the reader is referred to *Biopharmaceutical Drug Design and Development* (S. Wu-Pong et al. eds, Humana Press 1999); *Biopharmaceuticals: Biochemistry and Biotechnology* (G. Walsh, John Wiley & Sons, 1998); and the most current edition of Remington: *The Science and Practice of Pharmacy* (A. Gennaro, Lippincott, Williams & Wilkins). Pharmaceutical compositions of this invention may be packaged in a container with written instructions for use of the cells in human therapy, and the treatment of cancer.

25

5

10

15

20

The examples that follow are provided by way of further illustration, and are not meant to limit the claimed invention.

EXAMPLES

30

35

40

Example 1: Preparation of vectors controlling transcription in cells expressing telomerase reverse transcriptase

The lambda clone designated λGΦ5 containing the hTERT promoter is deposited with the American Type Culture Collection (ATCC), 10801 University Blvd., Manassas, VI 20110 U.S.A., under Accession No. 98505. λGΦ5 contains a 15.3 kbp insert including approximately 13,500 bases upstream from the hTERT coding sequence.

A Not1 fragment containing the hTERT promoter sequences was subcloned into the Not1 site of pUC derived plasmid, which was designated pGRN142. A subclone (plasmid pGRN140) containing a 9 kb Ncol fragment (with hTERT gene sequence and about 4 to 5 kb of lambda vector sequence) was partially sequenced to determine the orientation of the insert. pGRN140 was digested using Sall to

remove lambda vector sequences, the resulting plasmid (with removed lambda sequences) designated pGRN144. The pGRN144 insert was then sequenced.

SEQ. ID NO:1 is a listing of the sequence data obtained. Nucleotides 1-43 and 15376-15418 are plasmid sequence. Thus, the genomic insert begins at residue 44 and ends at residue 15375. The beginning of the cloned cDNA fragment corresponds to residue 13490. There are Alu sequence elements located ~1700 base pairs upstream. The sequence of the hTERT insert of pGRN142 can now be obtained from GenBank (http://www.ncbi.nlm.nih.gov/) under Accession PGRN142.INS AF121948. Numbering of hTERT residues for plasmids in the following description begins from the translation initiation codon, according to standard practice in the field. The hTERT ATG codon (the translation initiation site) begins at residue 13545 of SEQ. ID NO:1. Thus, position –1, the first upstream residue, corresponds to nucleotide 13544 in SEQ. ID NO:1.

5

10

15

20

25

30

35

40

Expression studies were conducted with reporter constructs comprising various hTERT upstream and intron sequences. A BgIII-Eco47III fragment from pGRN144 (described above) was digested and cloned into the BgIII-Nrul site of pSEAP2Basic (ClonTech, San Diego, CA) to produce plasmid designated pGRN148. A second reporter-promoter, plasmid pGRN150 was made by inserting the BgIII-FspI fragment from pGRN144 into the BgIII-Nrul sites of pSEAP2. Plasmid pGRN173 was constructed by using the EcoRV-Stul (from +445 to -2482) fragment from pGRN144. This makes a promoter reporter plasmid that contains the promoter region of hTERT from approximately 2.5 kb upstream from the start of the hTERT open reading frame to just after the first intron within the coding region, with the initiating Met codon of the hTERT open reading frame changed to Leu. Plasmid pGRN175 was made by APA1(Klenow blunt)-SRF1 digestion and religation of pGRN150 to delete most of the Genomic sequence upstream of hTERT. This makes a promoter/reporter plasmid that uses 204 nucleotides of hTERT upstream sequences (from position -36 to -117). Plasmid pGRN176 was made by PML1-SRF1 religation of pGRN150 to delete most of the hTERT upstream sequences. This makes a promoter/reporter plasmid that uses 204 nucleotides of hTERT upstream sequences (from position -36 to -239).

Levels of secreted placental alkaline phosphatase (SEAP) activity were detected using the chemiluminescent substrate CSPDTM (ClonTech). SEAP activity detected in the culture medium was found to be directly proportional to changes in intracellular concentrations of SEAP mRNA. The pGRN148 and pGRN150 plasmids (hTERT promoter-reporter) and the pSEAP2 plasmid (positive control, containing the SV40 early promoter and enhancer) were transfected into test cell lines. pGRN148 and pGRN150 constructs drove SEAP expression as efficiently as the pSEAP2 in immortal (tumor-derived) cell lines. Only the pSEAP2 control gave detectable activity in mortal cells.

The ability of the hTERT promoter to specifically drive the expression of the thymidine kinase (*tk*) gene in tumor cells was tested using a variety of constructs: One construct, designated pGRN266, contains an EcoRI-Fsel PCR fragment with the *tk* gene cloned into the EcoRI-Fsel sites of pGRN263. pGRN263, containing approximately 2.5 kb of hTERT promoter sequence, is similar to pGRN150, but contains a neomycin gene as selection marker. pGRN267 contains an EcoRI-Fsel PCR fragment with the *tk* gene cloned into the EcoRI-Fsel sites of pGRN264, pGRN264, containing approximately 210 bp of hTERT promoter sequence, is similar to pGRN176, but contains a neomycin gene as selection marker. pGRN268 contains an EcoRI-Xbal PCR fragment with the *tk* gene cloned into the EcoRI-Xbal

(unmethylated) sites of pGRN265. pGRN265, containing approximately 90 bp of hTERT promoter sequence, is similar to pGRN175, but contains a neomycin gene as selection marker.

These hTERT promoter/tk constructs, pGRN266, pGRN267 and pGRN268, were re-introduced into mammalian cells and tk/+ stable clones (and/or mass populations) were selected. Ganciclovir treatment in vitro of the tk/+ cells resulted in selective destruction of all tumor lines tested, including 143B, 293, HT1080, Bxpc-3', DAOY and NIH3T3. Ganciclovir treatment had no effect on normal BJ cells.

Figure 1 is a map of the TPAC adenovector pGRN376. It was made by cloning the NOT1-BAMH1 fragment from pGRN267 into the NOT1-BGL2 sites of pAdBN (Quantum Biotech). The 7185 bp vector comprises the herpes simplex thymidine kinase (tk) gene under control of the medium-length hTERT promoter sequence.

Example 2: Killing cancer cells using vectors controlled by the TERT promoter

5

10

15

20

25

.30

35

40

A replication-conditional adenovirus was constructed by placing a gene involved in viral replication under control of the hTERT promoter, which should activate transcription in telomerase-expressing cancer cells. The viral construct comprised the Inverted Terminal Repeat (ITR) from adenovirus Ad2; followed by the hTERT medium-length promoter (phTERT176) operably linked to the adenovirus E1a region; followed by the rest of the adenovirus deleted for the E3 region (Δ E3). As a positive control, a similar construct was made in which E1a was placed under control of the CMV promoter, which should activate transcription in any cell.

Reagents were obtained as follows. pBR322, restriction enzymes: NEB, Beverly, MA. Adenovirus Type 2 (Ad2), tissue culture reagents: Gibco/BRL, Grand Island, NY. Profection Mammalian Transfection Systems: Promega, Madison, WI. Tumor and Normal Cell lines: ATCC, Manassas, VA, except BJ line, which was obtained from J. Smith, U. of Texas Southwestern Medical Center.

Briefly, a pBR322-based plasmid was constructed which contains the Adenovirus Type 2 genome with deletions from 356-548nt (E1a promoter region) and 27971-30937nt (E3). A multiple cloning region was inserted at the point of deletion of the E1a promoter, and hTERT promoter (-239 to -36nt) or CMV promoter (-524 to -9nt) was subsequently cloned. Numbering of the CMV sequence is in accordance with Akrigg et al., Virus Res. 2:107, 1985. Numbering of the Ad2 sequence is in accordance with "DNA Tumor Viruses: Molecular Biology of Tumor Viruses", J. Tooze ed., Cold Spring Harbor Laboratory, NY.

These plasmid DNAs were digested with SnaBl to liberate ITRs, then phenol-chloroform extracted, precipitated and transfected into 293A cells for propagation of the virus. Several rounds of plaque purifications were performed using A549 cells, and a final isolate was expanded on these same cells. Viruses were titered by plaque assay on 293A cells, and tested for the presence of 5' WT Ad sequences by PCR. DNA was isolated from viruses by HIRT extraction.

Figure 2 shows the effect of these viruses on normal and cancer-derived cell lines. Each cell line was plated and infected at an MOI=20, ~24h post plating. The cells were then cultured over a period of 17-48 days, and fed every fourth day. The pictures shown in the Figure were taken 7 days after infection. The top row of each section shows the results of cells that were not virally infected (negative control). The middle row shows the results of cells infected with oncolytic adenovirus, in which replication gene E1a is operably linked to the hTERT promoter. The bottom row of each section shows the results of cells

infected with adenovirus in which E1a is operably linked to the CMV promoter (positive control). Results are summarized in Table 1.

TABLE 1: Effect of Oncolytic Virus on Cancerous and Non-cancerous Cells

| Cell Line | Origin | Culture Conditions | Uninfected cell Lysis | Lysis by phTERT- E1ΔE3 | Lysis by pCMV- E1ΔE3 |
|-----------|-----------------------------|--|-----------------------------|---------------------------|----------------------------|
| ВЈ | foreskin fibroblast | 90% DMEM/M199 + 10% FBS | NO | NO | YES |
| IMR | lung fibroblast | 90% DMEM/M199 + 10% FBS | NO | NO | YES |
| WI-38 | lung fibroblast | 90% DMEM/M199 + 10% FBS + 5 μg mL gentamicin | МО | NO | YES |
| A549 | lung carcinoma | 90% RPMI + 10% FBS | NO | YES | YES |
| AsPC-1 | adenocarcinoma, pancreas | 90% RPMI + 10% FBS | NO | YES | YES |
| BxPC-3 | adenocarcinoma, pancreas | 90% EMEM + 10% FBS | NO | YES | YES |
| DAOY | medulloblastoma | 90% EMEM + 10% FBS | NO | YES | YES |
| HeLa: | cervical carcinoma | 90% EMEM + 10% FBS | NO | YES | YES |
| HT1080 | fibrosarcoma | 90% EMEM + 10% FBS | NO | YES | YES |

5

10

15

20

All cell lines tested were efficiently lysed by AdCMV-E1dlE3 by day 17 post-infection. All tumor lines were lysed by AdphTERT-E1dlE3 in a similar, but slightly delayed period, while normal lines showed no signs of cytopathic effect and remained healthy out to 6 weeks post-infection.

The results demonstrate that an oncolytic virus can be constructed by placing a genetic element essential for replication of the virus under control of an hTERT promoter. Replication and lysis occurs in cancer cells, but not in differentiated non-malignant cells.

Example 3: Killing cancer cells using glycosyltransferase vectors and natural antibody

Adenovirus vectors comprising encoding sequences for glycosyltransferase under control of the TERT promoter are constructed by cloning the encoding sequence behind the hTERT promoter sequence of pGRN267, as described in Example 1.

SEQ. ID NO:2 and SEQ. ID NO:4 provide the encoding sequences for the A and B transferase, respectively.

Figure 3 is a comparison of the known mammalian α 1,3GT protein sequences, the ABO transferases, and the amino acid translation of the human α 1,3GT pseudogene. Based on this comparison and a comparison of the gene sequences, a humanized version of the marmoset α 1,3GT protein sequence has been devised (SEQ. ID NO:13). Another α 1,3GT sequence has been devised in which the marmoset prototype has been adapted with substitutions in the extracellular domain to enhance activity, based on a consensus of other mammalian α 1,3GT amino acid sequences (SEQ. ID NO:12).

Figure 4 provides a listing of a humanized α 1,3GT encoding sequence, adapting the marmoset nucleic acid sequence with conservative and silent substitutions in the human pseudogene (SEQ. ID NO:16).

A model adenovirus vector is made using the sheep α 1,3GT encoding sequence shown in SEQ. ID NO:17. Briefly, a Ecl136II fragment from a plasmid comprising the cDNA coding sequence plus ~70 bp of untranslated upstream sequence is cloned into the EcoRI(Klenow blunted)-Fsel(Klenow blunted) sites of pGRN267 such that the sheep α 1,3GT gene is in the same orientation as the hTERT promoter. Then a Notl-BamHI fragment from the plasmid containing the transcription pause region, the hTERT promoter, the sheep α 1,3GT gene sequence and the SV40 polyA signal is cloned into the Notl-BgIII sites of pAdBN (Quantum), which is then made into an adenovirus vector according to the manufacturer's technology.

5

10

15

20

Ability of α 1,3GT and ABO transferase vectors to promote tumor cell lysis is tested using a panel of established cell lines as in Example 2.

First, the ABO phenotype of each line is determined by incubating alternate wells with fresh human serum of the A and B blood type at 37°C for 30-60 min, and measuring trypan blue exclusion.

Fresh cells are then transduced with the test vectors at a suitable MOI, and cultured in a serum-free medium. Vectors comprising the opposite ABO transferase or α1,3GT under control of the TERT promoter are used to treat the test well. The same transferase under control of the CMV promoter is a positive control. A promoterless vector, a vector comprising ABO matched transferase, and empty vector can all serve as negative controls.

After 2 or 7 days, the cells are washed, and overlaid with fresh ABO matched human serum. After incubation at 37°C for 30-60 min, 0.4% trypan blue is added, and the percentage of lysed (blue staining) cells is determined.

SEQUENCE DATA

| | TABLE 2: Sequences listed in this Disclosure | | | | | |
|-------------|---|--|--|--|--|--|
| SEQ. ID NO: | Designation | Reference | | | | |
| 1 | Lambda clone designated λGφ5 (ATCC Accession No. 98505) Contains human Telomerase Reverse Transcriptase (hTERT) genomic insert (residues 44-15375). The ATG translation initiation site begins at residue 13545. | GenBank Accession AF121948 International Patent Publication WO 00/46355. | | | | |
| 2 | Human histo blood group A transferase cDNA sequence | GenBank Accession J05175 See also Accession Nos. AF134413 & AF134412; Yamamoto et al., Nature 1990 May 17;345:229 (1990); U.S. Patent 5,326,857 | | | | |
| 3 | Human histo blood group A transferase amino acid sequence | <i>(supra)</i> Figure 3 | | | | |
| 4 | Human histo blood group B transferase cDNA sequence | GenBank Accession AF134414 Yamamoto et al., Nature 1990 May 17;345:229 (1990); U.S. Patent 5,326,857 | | | | |
| 5 | Human histo blood group B transferase amino acid sequence | <i>(supra)</i> Figure 3 | | | | |
| 6 | Marmoset α1,3-galactosyltransferase amino acid sequence | GenBank Accession S71333 Henion et al., Glycobiology 4,193 (1994) Figure 3 | | | | |
| 7 | Amino acid translation of human 1,3-galactosyltransferase pseudogene | <i>(infra)</i> Figure 3 | | | | |
| 8 | Sheep α1,3-galactosyltransferase amino acid sequence | Chris Denning & John Clark, Geron Biomed Figure 3 | | | | |
| 9 | Bovine α1,3-galactosyltransferase amino acid sequence | GenBank Accession J04989 Joziasse et al. "Bovine α1->3- galactosyltransferase" J. Biol. Chem. 264, 14290 (1989) Figure 3 | | | | |
| 10 | Pig α1,3-galactosyltransferase amino acid sequence | GenBank Accession L36152 Sus scrofa alpha-1,3-galactosyltransferase mRNA. Strahan et al. "cDNA sequence and chromosome localization of pig α1,3 galactosyltransferase" Immunogenetics 41, 101 (1995) See also GenBank Accession L36535 Sandrin et al. "Characterization of cDNA clones for porcine a(1,3)galactosyl transferase" Xenotransplantation (1994) Figure 3 | | | | |

17

Chris Denning & John Clark, Geron Biomed

| TABLE 2: Sequences listed in this Disclosure | | | | | |
|--|--|--|--|--|--|
| SEQ. ID NO: | Designation | Reference | | | |
| 11 | Mouse α1,3-galactosyltransferase amino acid sequence | GenBank Accession M26925 Larsen et al. "Isolation of a cDNA encoding a murine UDPgalactose:β-D-galactosyl-1,4-N-acetyl-D-glucosaminide alpha-1,3-galactosyltransferase" Proc. Natl. Acad. Sci. USA 86, 8227 (1989) See also GenBank Accession IM85153 Joziasse et al. "Murine alpha-1,3-galactosyltransferase: A single gene locus specifies four isoforms of the enzyme by alternative splicing" J. Biol. Chem. 267, 5534 (1992) Figure 3 | | | |
| 12 | Consensus α1,3-galactosyltransferase amino acid sequence | This Invention Figure 3 | | | |
| 13 | Humanized α 1,3-galactosyltransferase amino acid sequence | This Invention Figure 3 | | | |
| 14 | Marmoset α1,3-galactosyltransferase cDNA sequence | GenBank Accession S71333 Henion et al., Glycobiology 4,193 (1994) Figure 4 | | | |
| 15 | Human α1,3-galactosyltransferase pseudogene sequence | GenBank Accession J05421 Larsen et al., J. Biol. Chem265:7055, 1990 See also GenBank Accession M60263 Joziasse et al. "Characterization of an alpha-1->3-galactosyltransferase homologue on human chromosome 12 that is organized as a processed pseudogene" J. Biol. Chem. 266, 6991 (1991) Figure 4 | | | |
| 16 | Humanized α 1,3-galactosyltransferase encoding sequence | This Invention Figure 4 | | | |

SEQ. ID NO:1 (human TERT promoter & upstream sequence)

Sheep α 1,3-galactosyltransferase encoding sequence

| gcggccgcga | gctctaatac | gactcactat | agggcgtcga | ctcgatcaat | ggaagatgag | 60 |
|------------|------------|------------|------------|------------|------------|------|
| gcattgccga | agaaaagatt | aatggatttg | aacacacagc | aacagaaact | acatgaagtg | 120 |
| aaacacagga | aaaaaaagat | aaagaaacga | aaagaaaagg | gcatcagtga | gcttcagcag | 180 |
| aagttccatc | ggccttacat | atgtgtaagc | agaggccctg | taggagcaga | ggcaggggga | 240 |
| aaatacttta | agaaataatg | tctaaaagtt | tttcaaatat | gaggaaaaac | ataaaaccac | 300 |
| agatccaaga | agctcaacaa | aacaaagcac | aagaaacagg | aagaaattaa | aagttatatc | 360 |
| acagtcaaat | tgctgaaaac | cagcaacaaa | gagaatatct | taagagtatc | agaggaaaag | 420 |
| agattaatga | caggccaaga | aacaatgaaa | acaatacaga | tttcttgtag | gaaacacaag | 480 |
| acaaaagaca | ttttttaaaa | ccaaaaggaa | aaaaaatgct | acattaaaat | gttttttacc | 540 |
| cactgaaagt | atatttcaaa | acatatttta | ggccaggctt | ggtggctcac | acctgtaatc | 600 |
| ccagcacttt | gggaggccaa | ggtgggtgga | tcgcttaagg | tcaggagttc | gagaccagcc | 660 |
| tggccaatat | agcgaaaccc | catctgtact | aaaaacacaa | aaattagctg | ggtgtggtga | 720 |
| cacatgcctg | taatcccagg | tactcaggag | gctaaggcag | gagaattgct | tgaactggga | 780 |
| ggcagaggtg | gtgagccaag | attgcaccag | tgcactccag | ccttggtgac | agagtgaaac | 840 |
| tccatctcaa | aaacaaacaa | acaaaataca | tatacataaa | tatatatgca | catatatata | 900 |
| catatataaa | tatatataca | catatataaa | tctatataca | tatatacata | tatacacata | 960 |
| tataaatcta | tatacatata | tatacatata | taatatattt | acatatataa | atatatacat | 1020 |
| atataaatat | acatatataa | atacatatat | aaatatacat | atataaatat | acatatataa | 1080 |
| atatacatat | ataaatatat | acatatataa | atatacatat | ataaatatat | atacatatat | 1140 |
| aaatatataa | atatacaagt | atatacaaat | atatacatat | ataaatgtat | atacgtatat | 1200 |
| acatatatat | ataaatatat | aaaaaaactt | ttggctgggc | acctttccaa | atctcatggc | 1260 |
| acatataagt | ctcatggtaa | cctcaaataa | aaaaacatat | aacagataca | ccaaaaataa | 1320 |
| aaaccaataa | attaaatcat | gccaccagaa | gaaattacct | tcactaaaag | gaacacagga | 1380 |

| aggaaagaaa ggagtaattc | | | | | | 7 4 4 0 |
|--|---|---|--|---|---|--|
| | yaayyaayay | aagaccatga | aacaaccaga | aaacaaacaa | caaaacagca | 1440 |
| ggagtaatte | | | | | | 1500 |
| | | | | | | |
| aaaagacata | gagtggctga | atggacgaaa | aaaacaagac | tcaataatct | gttgcctaca | 1560 |
| | tcacctataa | | | | | 1620 |
| | | | | | | |
| ctatgcaaat | ggaaaccaaa | aaaagaacag | aactagctac | acttatatca | gacaaaatag | 1680 |
| | aaaaagtaca | | | | | 1740 |
| | | | | | | |
| tataacaatt | gtgaatttat | atgcgcccaa | cactgggaca | cccagatata | tacagcaaat | 1800 |
| | ctaaggagag | | | | | 1860 |
| | | | | | | |
| gcttttagca | ttggacagat | catccagaca | gaaaatcaac | caaaaaattg | gacttaatct | 1920 |
| atsatataga | acaaatgtac | ctaattoato | tttacaanac | atttratcra | ntanttocan | 1980 |
| | | | | | | - |
| aatatgcatt | ttttcctcag | catatggatc | attctcaagg | atagaccata | tattaggcca | 2040 |
| canaacaanc | cattaaaaat | traassaat | tranccarac | atdatoactt | atocttotaa | 2100 |
| | | | | | | |
| ttacagcact | ttggggaggg | τgaggτggga | ggatgtcttg | agtacaggag | tttgagacca | 2160 |
| acctagacaa | aatagtgaga | ccctatctct | acaaactttt | tttttttaatt | agccaggcat | 2220 |
| | | | | | | 2280 |
| | gcctgtagtc | | | | | |
| ccaagagttc | aaggctacgg | toaoccatoa | ttgcaacacc | acacaccagc | cttootoaca | 2340 |
| | | | | | | 2400 |
| gaatgagacc | ctgtctcaaa | adddadadaa | aaaattgaaa | taatataaay | Cattlette | |
| ggccacagtg | gaacaaaacc | agaaatcaac | aacaagagga | attttqaaaa | ctatacaaac | 2460 |
| | | | | | | 2520 |
| | taaacaatat | | | | | |
| gaaattgaaa | aatttattta | aqcaaatqat | aacqqaaaca | taacctctca | aaacccacgg | 2580 |
| | | | | | | 2640 |
| | aagcagtgct | | | | | |
| gtagaaaagc | caggcgcagt | ggctcatgcc | tgtaatccca | gcactttggg | aggccaaggc | 2700 |
| | cctgaggtca | | | | | 2760 |
| | | | | | | |
| cgctactaaa | aatacaaaat | tagctgggca | tggtggcaca | tgcctgtaat | cccagctact | 2820 |
| | aggcaggata | | | | | 2880 |
| | | | | | | |
| attgcgccat | tggactccag | cctgggtaac | aagagtgaaa | ccctgtctca | agaaaaaaaa | 2940 |
| senstnesse | aaacttaaaa | atacaaccta | atoatocacc | ttaaanaact | 2022220022 | 3000 |
| aaaag tagaa | uuuc c cuuuu | d cacaacc ca | acguigeace | teaagaacc | | |
| gagcaaacta | aacctaaaat | τggτaaaaga | aaagaaataa | taaagatcag | agcagaaata | 3060 |
| | aaagataaca | | | | | 3120 |
| | | | | | | |
| agataaacaa | aattgacaaa | cctttgccca | gactaagaaa | aaaggaaaga | agacctaaat | 3180 |
| aaataaantr | agagatgaaa | aaananacat | tacaactgat | accacagaaa | ttcaaaggat | 3240 |
| | | | | | | 3300 |
| cactagaggc | tactatgagc | aactgtacac | taataaattg | aaaaacctag | aaaaaalaga | |
| taaatteeta | gatgcataca | acctaccaag | attgaaccat | gaagaaatcc | aaadcccaaa | 3360 |
| | | | | | | 3420 |
| cagaccaata | acaataatgg | gattaaagcc | ataataaaaa | gtctcctagc | aaagagaagc | |
| ccannaccca | atggcttccc | toctogattt | taccaatcat | ttaaagaaga | atgaattcca | 3480 |
| | | | | ****** | attetaeste | 3540 |
| | aactattctg | | | | | |
| accaatatta | ccctgattcc | aaaaccagac | aaaaacacat | caaaaacaaa | caaacaaaaa | 3600 |
| | | ******** | tecetantan | atacteatec | 22222tccte | 3660 |
| aacagaaaga | aagaaaacta | cayyccaata | tecetyatya | atactyatac | aaaaatttt | |
| aacaaaacac | tagcaaacca | aattaaacaa | caccttcgaa | agatcattca | ttgtgatcaa | 3720 |
| ******** | tteeseast | aannaaataa | ****** | ncanatenat | castatasta | 3780 |
| gigggailla | ttccagggat | yyaayyatyy | licadialai | ycaaaccaac | caatytyata | |
| catcatccca | acaaaatgaa | qtacaaaaac | tatatgatta | tttcacttta | tgcagaaaaa | 3840 |
| | aaattctgca | | | | | 3900 |
| | | | | | | |
| aaacatacag | gccaggcaca | gtggctcaca | cctgcgatcc | cagcactctg | ggaggccaag | 3960 |
| | tgcttgggcc | | | | | 4020 |
| | | | | | | |
| gtctacaaaa | aactttttta | aaaaattagc | caggcatgat | ggcatatgcc | tgtagtccca | 4080 |
| actanteton | aggctgaggt | addanaatca | cttaagccta | ggaggtcgag | actacaataa | 4140 |
| | | | | | | |
| gccatgaaca | tgtcactgta | ctccagccta | gacaacagaa | caagacccca | ctgaataaga | 4200 |
| | | | | aggaggagaa | ntancancen | |
| agaaggagaa | ggagaaggga | gaaaggaggg | agaagggagg | | | 4260 |
| | ggagaaggga | | | | | |
| | | | | | | 4320 |
| gaggagaagt | ggaaggggaa | ggggaaggga | aagaggaaga | agaagaaaca | tatttcaaca | 4320 |
| gaggagaagt taataaaagc | ggaaggggaa cctatatgac | ggggaaggga agaccgaggt | aagaggaaga agtattatga | agaagaaaca ggaaaaactg | tatttcaaca aaagcctttc | 4320 4380 |
| gaggagaagt taataaaagc | ggaaggggaa cctatatgac | ggggaaggga agaccgaggt | aagaggaaga agtattatga | agaagaaaca ggaaaaactg | tatttcaaca aaagcctttc | 4320 4380 4440 |
| gaggagaagt taataaaagc ctctaagatc | ggaaggggaa cctatatgac tggaaaatga | ggggaaggga agaccgaggt caagggccca | aagaggaaga agtattatga ctttcaccac | agaagaaaca ggaaaaaactg tgtgattcaa | tatttcaaca aaagcctttc catagtacta | 4320 4380 4440 |
| gaggagaagt taataaaagc ctctaagatc gaagtcctag | ggaaggggaa cctatatgac tggaaaatga ctagagcaat | ggggaaggga agaccgaggt caagggccca cagataagag | aagaggaaga agtattatga ctttcaccac aaagaaataa | agaagaaaca ggaaaaactg tgtgattcaa aaggcatcca | tatttcaaca aaagcctttc catagtacta aactggaaag | 4320 4380 4440 4500 |
| gaggagaagt taataaaagc ctctaagatc gaagtcctag gaagaagtca | ggaaggggaa cctatatgac tggaaaatga ctagagcaat aattatcctg | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt | agaagaaaca ggaaaaactg tgtgattcaa aaggcatcca atatctggaa | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag | 4320 4380 4440 4500 4560 |
| gaggagaagt taataaaagc ctctaagatc gaagtcctag gaagaagtca | ggaaggggaa cctatatgac tggaaaatga ctagagcaat aattatcctg | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt | agaagaaaca ggaaaaactg tgtgattcaa aaggcatcca atatctggaa | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag | 4320 4380 4440 4500 |
| gaggagaagt taataaaagc ctctaagatc gaagtcctag gaagaagtca acaccactaa | ggaaggggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc | agaagaaaca ggaaaaactg tgtgattcaa aaggcatcca atatctggaa aggatacaaa | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac | 4320 4380 4440 4500 4560 4620 |
| gaggagaagt taataaaagc ctctaagatc gaagtcctag gaagaagtca acaccactaa aaaaatcagt | ggaaggggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta agtatttcta | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc | agaagaaaca ggaaaaactg tgtgattcaa aaggcatcca atatctggaa aggatacaaa tgaaaaagaa | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag | 4320 4380 4440 4500 4560 4620 4680 |
| gaggagaagt taataaaagc ctctaagatc gaagtcctag gaagaagtca acaccactaa aaaaatcagt | ggaaggggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta agtatttcta | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc | agaagaaaca ggaaaaactg tgtgattcaa aaggcatcca atatctggaa aggatacaaa tgaaaaagaa | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag | 4320 4380 4440 4500 4560 4620 |
| gaggagaagt taataaaagc ctctaagatc gaagtcctag gaagaagtca acaccactaa aaaaatcagt cagctacaaa | ggaaggggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta agtatttcta taaaattaaa | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag | agaagaaaca ggaaaaactg tgtgattcaa aaggcatcca atatctggaa aggatacaaa tgaaaaagaa aagtgaaaga | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tctctacaat | 4320 4380 4440 4500 4560 4620 4680 4740 |
| gaggagaagt taataaaagc ctctaagatc gaagtcctag gaagaagtca acaccactaa aaaaatcagt cagctacaaa gaaaactata | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta agtattcta taaaattaaa aaatattgat | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagata gagctgaaat tattccaaca cagctaggaa aaaagaaatt | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca | agaagaaaca ggaaaaactg tgtgattcaa aaggcatcca atatctggaa aggatacaaa tgaaaagaa aagtgaaaga caaaaaaga | tattcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaaag tctctacaat aaagatattc | 4320 4380 4440 4500 4560 4620 4680 4740 4800 |
| gaggagaagt taataaaagc ctctaagatc gaagtcctag gaagaagtca acaccactaa aaaaatcagt cagctacaaa gaaaactata catgttcata | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta agtatttcta taaaaattaaa aaatattgat gattggaaga | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt ataaatactg | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaaaatgtc | agaagaaca ggaaaaactg tgtgattcaa aaggcatcca atatctggaa aggatacaaa tgaaaaagaa aagtgaaaga caaaaaaga caaaaaaga | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tcttacaat aaagatattc aaagcaattt | 4320 4380 4440 4500 4560 4620 4680 4740 4800 4860 |
| gaggagaagt taataaaagc ctctaagatc gaagtcctag gaagaagtca acaccactaa aaaaatcagt cagctacaaa gaaaactata catgttcata | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta agtatttcta taaaaattaaa aaatattgat gattggaaga | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt ataaatactg | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaaaatgtc | agaagaaca ggaaaaactg tgtgattcaa aaggcatcca atatctggaa aggatacaaa tgaaaaagaa aagtgaaaga caaaaaaga caaaaaaga | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tcttacaat aaagatattc aaagcaattt | 4320 4380 4440 4500 4560 4620 4680 4740 4800 |
| gaggagaagt taataaaagc ctctaagatc gaagtcctag gaagaagtca acaccactaa aaaaatcagt cagctacaaa gaaaactata catgttcata acaaattcaa | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta agtatttcta taaaattaaa gattggaaga tgcaatccct | ggggaaggga agaccgaggt caagggccca ttgcagata gagctgaaat tattccaaca cagctaggaa aaaagaaat ataaatactg attaaaatac | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaaccaaac gaagagggca ttaaaatgtc taatgacgtt | agaagaaca ggaaaaactg tgtgattcaa aaggcatcca atatctggaa aggatacaaa tgaaaaagaa caaaaaaaga catactaccc | tattcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tcttacaat aaagatattc aaagcaattt atagaagaaa | 4320 4380 4440 4500 4560 4620 4680 4740 4800 4860 4920 |
| gaggagaagt taataaaagc ctctaagatc gaagtactag gaagaagtca acaccactaa aaaaaactagt cagctacaaa gaaaactata catgttcata acaaattcaa caattctaag | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta agtatttcta taaaattaaa aaatattgat gattggaaga tgcaatccct atttgtacag | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt ataaatactg attaaaatac aaccaaaa | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaaaaatgtc taaatgacgtt gacccagaat | agaagaaca ggaaaaactg ggtgatctaa aaggcatcca atatctggaa aggatacaaa aggtgaaaga caaaaaagaa catactaccc cttcacagaa agccaaagct | tatttcaaca aaagcctttt catagtacta aactggaaag aagacttaag atcaatgtaa cacaaaaaag tctctacaat aaagatattc aaagcaattt atagaagaaa atcctgacca | 4320 4380 4440 4500 4560 4620 4680 4740 4800 4860 4920 4980 |
| gaggagaagt taataaaagc ctctaagatc gaagtactag gaagaagtca acaccactaa aaaaaactagt cagctacaaa gaaaactata catgttcata acaaattcaa caattctaag | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta agtatttcta taaaattaaa aaatattgat gattggaaga tgcaatccct atttgtacag | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt ataaatactg attaaaatac aaccaaaa | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaaaaatgtc taaatgacgtt gacccagaat | agaagaaca ggaaaaactg ggtgatctaa aaggcatcca atatctggaa aggatacaaa aggtgaaaga caaaaaagaa catactaccc cttcacagaa agccaaagct | tatttcaaca aaagcctttt catagtacta aactggaaag aagacttaag atcaatgtaa cacaaaaaag tctctacaat aaagatattc aaagcaattt atagaagaaa atcctgacca | 4320 4380 4440 4500 4560 4620 4680 4740 4800 4860 4920 |
| gaggagaagt taataaaagc ctctaagatc gaagtcctag gaagaagtca acaccactaa aaaaatcagt cagctacaaa gaaaactata catgttcata acaaattcaa acaattcaag aaaagaacaa | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta agtatttcta taaaattaaa aaatattgat gattggaaga tgcaatccct atttgtacag aactggaagc | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt attaaaatac attaaaatac aaccacaaaa atcacattac | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaaaatgtc taatgacgtt gacccagaat ctgacttcaa | agaagaaca ggaaaaactg tgtgattcaa aaggcatcca atatctggaa aggatacaaa tgaaaaagaa aagtgaaaga caaaaaaaga catactaccc cttcacagaa agccaaagct attatactac | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tctctacaat aaagatattc aaagcaattt atagaagaaa atcctgacca aaagctatag | 4320 4380 4440 4500 4560 4620 4680 4740 4800 4860 4920 4980 5040 |
| gaggagaagt taataaaagc ctctaagatc gaagtcctag gaagaagtca acaccactaa aaaaatcagt cagctacaaa gaaaactata catgttcata acaaattcaa caattctaag aaaagaacaa taacccaaac | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaactatta agtatttcta ataaaattaaa aaatattgat gattggaaga tgcaatccct atttgtacag aactggaagc tacatggaagc | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt ataaatactg attaaaatac aaccacaaaa atcacattac tggcataaaa | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaaaatgtc taatgacgtt gacccagaat ctgacttcaa acagatgaga | agaagaaaca ggaaaaactg tgtgattcaa aaggcatcca atatctggaa aggatacaaa tgaaaaagaa caaaaaaaga catactaccc cttcacagaa agccaaagct attatactac catggaccag | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tctctacaat aaagatattc aaagcaattt atagaagaaa atcctgacca aagctatag aggaacagaa | 4320 4380 4440 4500 4560 4620 4680 4740 4800 4860 4920 4980 5040 5100 |
| gaggagaagt taataaaagc ctctaagatc gaagtcctag gaagaagtca acaccactaa aaaaatcagt cagctacaaa gaaaactata catgttcata acaaattcaa caattctaag aaaagaacaa taacccaaac | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaactatta agtatttcta ataaaattaaa aaatattgat gattggaaga tgcaatccct atttgtacag aactggaagc tacatggaagc | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt ataaatactg attaaaatac aaccacaaaa atcacattac tggcataaaa | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaaaatgtc taatgacgtt gacccagaat ctgacttcaa acagatgaga | agaagaaaca ggaaaaactg tgtgattcaa aaggcatcca atatctggaa aggatacaaa tgaaaaagaa caaaaaaaga catactaccc cttcacagaa agccaaagct attatactac catggaccag | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tctctacaat aaagatattc aaagcaattt atagaagaaa atcctgacca aagctatag aggaacagaa | 4320 4380 4440 4500 4560 4620 4680 4740 4800 4860 4920 4980 5040 |
| gaggagaagt taataaaagc ctctaaggatcgaagtcctag gaagaagtca acaccactaa aaaaatcagt cagctacaaa gaaaactata acatgttcata acaaattcaa caattctaag aaaagaacaa tagacaaac tagagaatcc | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta agtatttcta taaaattaaa aaatattgat gattggaaga tgcaatccct atttgtacag aactggaagc agaacaagata | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt ataaatacta atcacataca atcacataca cagcataaaa ccatgcatca | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaaaaatgtc taatgacgtt gacccagaat ctgacttcaa acagatgaga acagtgagaa | agaagaaaca ggaaaaacta ggaatcaa aaggcatcca atatctggaa aggatacaaa tgaaaagaa aagtgaaaga caaaaaaga catactacc cttcacagaa agccaaagct attatactac attatgaccag cattttgac | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tctctacaat aaagatattc atagaagaaa atcctgacca aaagctatag aggaacagaa aaaggtgcca | 4320 4380 4440 4500 4560 4620 4680 4740 4860 4920 4980 5040 5160 |
| gaggagaagt taataaaagc ctctaagatc gaagtcctag gaagaagtca acaccactaa aaaaacagt cagctacaaa gaaaactata catgttcata acaattctaag aaaagaacaa taacccaaac tagagaatcc aggagaatcc | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta agtatttcta taaaattaaa aaatattgat gattggaaga tgcaatccct atttgtacag aactggaagc tacatggtac agaaacaaat ttggggaaaa | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt ataaatactg attaaaatac aaccacaaaa atcacattac tggcataaaa ccatgcatct gataatct | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaaaatgtc taatgacgtt gaccagaat ctgacttcaa acagatgaga acagtgaga tcaataaatg | agaagaaaca ggaaaaactg tgtgattcaa aaggcatcca atatctggaa aggatacaaa aagtgaaaga caaaaaagaa catactaccc cttcacagaa agccaaagct attatactac catggaccag catttttgac gtgctggagg | tatttcacca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tctctacaat taaagatattc aaagcaattt atagaagaaa atcctgacca aaagctatag aggaacagaa aaaggtgcca aactggatat | 4320 4380 4440 4500 4560 4620 4680 4740 4800 4920 4980 5040 5100 5160 5220 |
| gaggagaagt taataaaagc ctctaagatc gaagtcctag gaagaagtca acaccactaa aaaaacagt cagctacaaa gaaaactata catgttcata acaattctaag aaaagaacaa taacccaaac tagagaatcc aggagaatcc | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta agtatttcta taaaattaaa aaatattgat gattggaaga tgcaatccct atttgtacag aactggaagc tacatggtac agaaacaaat ttggggaaaa | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt ataaatactg attaaaatac aaccacaaaa atcacattac tggcataaaa ccatgcatct gataatct | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaaaatgtc taatgacgtt gaccagaat ctgacttcaa acagatgaga acagtgaga tcaataaatg | agaagaaaca ggaaaaactg tgtgattcaa aaggcatcca atatctggaa aggatacaaa aagtgaaaga caaaaaagaa catactaccc cttcacagaa agccaaagct attatactac catggaccag catttttgac gtgctggagg | tatttcacca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tctctacaat taaagatattc aaagcaattt atagaagaaa atcctgacca aaagctatag aggaacagaa aaaggtgcca aactggatat | 4320 4380 4440 4500 4560 4620 4680 4740 4800 4920 4980 5040 5100 5160 5220 |
| gaggagaagt taataaaagc ctctaagatc gaagtcctag gaagaagtca acaccactaa aaaaatcagt cagctacaaa gaaaactata catgttcata acaaattcaa aaaagaacaa taacccaaac tagagaatcc agagaatcc agagaatcc catatgcaa | ggaagggaa cctatatgac tggaaaatga ctagaaactatta aatatcctg aaaactatta agtatttcta taaaattaaa aaatattgat gattggaaga tgcaatccct aattgtacag aactggaagc tacatggtac agaaacaaa ttggggaaaa aataacaata | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt attaaaatac atcacacaaa atcacattac tggcataaaa ccatgcatct gataatctc tgataatctc | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaaaatgtc taatgacgtt gacccagaat ctgacttcaa acagatgaga acagtgaaga tcaataaatg gtctctcacc | agaagaaca ggaaaaactg tgtgattcaa aaggcatcca atatctggaa aggatacaaa tgaaaaagaa caaaaaaaga catactaccc cttcacagaa agccaaagct attatactac catggaccag cattittga cattittga gtgtggagg atatacaaaa | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tctctacaat aaagcaattt atagaagaaa atcctgacca aaagctatag aggaacagaa aaaggtgcca aaaggtgcca aactggatat gcaaatcaaa | 4320 4380 4440 4500 4560 4620 4680 4740 4860 4920 4980 5040 5160 5220 5280 |
| gaggagaagt taataaaagc ctctaagatc gaagtcctag gaagaagtca acaccactaa aaaaatcagt cagctacaaa gaaaactata caattctaa aaaattcaa aaaatcaaa taacccaaac tagagaatca tagagaatca agagaacata tcatatgcaaa tagagaacaa tagagaacaa atgagaacaaa tcatatgcaa atggatgaaa | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta agtatttcta agtatttcta gattggaaga tgcaatccct atttgtacag aactggaagc tacatggtaca gaacagaat aggaaacaat tgggggaaaa aggaacaat tgggggaaaa | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt ataaatactg attaaaatac aaccacaaaa atcacattac tggcataaac ccatgcatct gataatctc tctagaactct | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaaaatgtc taatgacgtt gacccagaat ctgacttcaa acagatgaga acagtgaact tcaataaatg gtctctcacc aactttgcaa | agaagaaaca ggaaaaactg atatctggaa aggatacaaa aggaaaagga catactaccc cttcacagaa agccaagct attatactac catggaccag cattittgac gtgctggagg atatacaaaa ctactaaaag ctactaaaag | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tctctacaat aaagcaattt atagaagaaa atcctgacca aaagctatag aggaacagaa aaaggtgcca acaggtgcca acctggatat gcaaatcaaa aaaacaccgg | 4320 4380 4440 4500 4560 4620 4680 4740 4860 4920 4980 5040 5100 5160 5220 5280 5340 |
| gaggagaagt taataaaagc ctctaagatc gaagtcctag gaagaagtca acaccactaa aaaaatcagt cagctacaaa gaaaactata caattctaa aaaattcaa aaaatcaaa taacccaaac tagagaatca tagagaatca agagaacata tcatatgcaaa tagagaacaa tagagaacaa atgagaacaaa tcatatgcaa atggatgaaa | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta agtatttcta agtatttcta gattggaaga tgcaatccct atttgtacag aactggaagc tacatggtaca gaacagaat aggaaacaat tgggggaaaa aggaacaat tgggggaaaa | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt ataaatactg attaaaatac aaccacaaaa atcacattac tggcataaac ccatgcatct gataatctc tctagaactct | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaaaatgtc taatgacgtt gacccagaat ctgacttcaa acagatgaga acagtgaact tcaataaatg gtctctcacc aactttgcaa | agaagaaaca ggaaaaactg atatctggaa aggatacaaa aggaaaagga catactaccc cttcacagaa agccaagct attatactac catggaccag cattittgac gtgctggagg atatacaaaa ctactaaaag ctactaaaag | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tctctacaat aaagcaattt atagaagaaa atcctgacca aaagctatag aggaacagaa aaaggtgcca acaggtgcca acctggatat gcaaatcaaa aaaacaccgg | 4320 4380 4440 4500 4560 4620 4680 4740 4860 4920 4980 5040 5160 5220 5280 |
| gaggagaagt taataaaagc catcaagatc gaagtcctag gaagaagtca acaccactaa aaaaatcagt cagctacaaa gaaaactata acaagttcaa caattctaag aaaagaacaa taagccaaac tagagaatcc agaacatact ccatatgcaa atggatgaaa aggaaacctc | ggaagggaa cctatatgac ctagagcaat aattatcctg aaaactatta agtattcta taaaattaaa aaatattgat gattggaaga tgcaatccct atttgtacag aactggaagc tacatggtac agaaacaaat ttggggaaaa agaaacaat tgggggaaaa | ggggaaggga agaccgaggg caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt attaaaatact attacaatac atgcataaa ccatgcatct gataatctct ctagaactct taaaacctca gagtgggcaa | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagg gcaaacaatc ttaaccaaag gaagagggca ttaaatagtc taatgacgtt gacccagaat ctgacttcaa acagttgaaa acagtgaact tcaataaatg gtctctcacc aactttgcaa agactttgaa | agaagaaaca ggaaaaacta ggaatacaa aggcatcca atatctggaa aggtacaaa tgaaaaagaa aagtgaaaga caaaaaaga catactaccc cttcacagaa agccaaagct attatactac catggaccag cattttgac gtgctggagg atatacaaaa ctactaaaag agtaattccc | tattcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac aacaaaaaag tctctacaat aaagatattc atagaagaaa atcctgacca aaagctataag aggaacagaa aaggacagaa aaaggtgcca aactggatat gcaaatcaaa aaacacccgg tgcaggcaca | 4320 4380 4440 4500 4560 4620 4680 4740 4860 4920 4980 5100 5160 5220 5280 5340 5400 |
| gaggagaagt taataaaagc ctctaagatc gaagtcctag gaagaagtca acaccactaa aaaaaccagt cagctacaaa gaaaactata catgttcata acaattctaag aaaagaacaa taacccaaac taggagaatcc tagaagaatcc cagaacatact ccatatgcaa atggatgaaa atggatgaaa atggatgaaa aggaaactctc ggcaaccaaa | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta agtatttcta taaaattaaa aaatattgat gattggaaga tgcaatccct atttgtacag aactggaagc tacatggtac agaaacaaat ttggggaaaa aataacaata ggcttaaatc caggacattg gcaaaacag | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt ataaatactg atcacaaaa atcacataac tggcataaaa ccatgcatct gataatctt taaaactct gataatctt taaaactca gagtgggcaa acaaatggga | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaaaatgtc taatgacgtt gacccagaat ctgacttcaa acagatgaga acagtgaga tcaataaatg gtctctcacc aactttgcaa agacttcttg tcataag | agaagaaaca ggaaaaactg tgtgattcaa aaggcatcca atactggaaaaga aagtgaaaga catactaccc cttcacagaa agccatattgac gtgctggagg atatacaaaa ctactaaaag agtaattccc ttaaaaagct taaaaagct | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tctctacaat ataagatattc aaagcaattt atagaagaaa atcctgacca aaagctatag aggaacagaa aaaggtgcca aactggatat gcaaatcaaa aaaaacaccgg tgcaggcaca tctgcccag | 4320 4380 4440 4560 4620 4620 4680 4740 4860 4980 5040 5100 5160 5220 5280 5340 5460 |
| gaggagaagt taataaaagc ctctaagatc gaagtcctag gaagaagtca acaccactaa aaaaaccagt cagctacaaa gaaaactata catgttcata acaattctaag aaaagaacaa taacccaaac taggagaatcc tagaagaatcc cagaacatact ccatatgcaa atggatgaaa atggatgaaa atggatgaaa aggaaactctc ggcaaccaaa | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta agtatttcta taaaattaaa aaatattgat gattggaaga tgcaatccct atttgtacag aactggaagc tacatggtac agaaacaaat ttggggaaaa aataacaata ggcttaaatc caggacattg gcaaaacag | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt ataaatactg atcacaaaa atcacataac tggcataaaa ccatgcatct gataatctt taaaactct gataatctt taaaactca gagtgggcaa acaaatggga | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaaaatgtc taatgacgtt gacccagaat ctgacttcaa acagatgaga acagtgaga tcaataaatg gtctctcacc aactttgcaa agacttcttg tcataag | agaagaaaca ggaaaaactg tgtgattcaa aaggcatcca atactggaaaaga aagtgaaaga catactaccc cttcacagaa agccatattgac gtgctggagg atatacaaaa ctactaaaag agtaattccc ttaaaaagct taaaaagct | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tctctacaat ataagatattc aaagcaattt atagaagaaa atcctgacca aaagctatag aggaacagaa aaaggtgcca aactggatat gcaaatcaaa aaaaacaccgg tgcaggcaca tctgcccag | 4320 4380 4440 4500 4560 4620 4680 4740 4860 4920 4980 5100 5160 5220 5280 5340 5400 |
| gaggagaagt taataaaagc ctctaaggatcgaagtcaaaaaaacaatcaa aaaaactata catgttcata acaaattcaag aaaagaacaa taacccaaac taggagaacca caggacatacc ccatatgcaa atggatgaaa aggaacatc ccgaaaccaaa | ggaagggaa cctatatgac tggaaaatga ctagagcaat aatatcctg aaaactatta agtatttcta taaaattaaa aaatattgat gattggaaga tgcaatccct aattgtacag aactggaagc tacatggtac agaaacaaa ttggggaaaa aataacaata ggcttaaatc caggacattg gcaaaacag actggaaacaaa | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt ataaaatacta atcacacaaaa atcacataac tggcataaaa ccatgcatct gataatctct ctagaactct taaaacccca gagtgggca agaagagaaca | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaaaatgtc taatgacgtt gacccagaat ctgacttcaa acagatgaga acagtgaac tcaataaaatg gtctctcacc aactttgcaa agacttcttg tcatatcaag acccacagaa | agaagaaaca ggaaaacatca aagcaaaagaa aagtgaaaga catactacca agcaaagaca agcaaagaca agcaaagca attatactac catggaccag attatacaaa agtaatacaaaa ctactaaaag agtaattcc tgggagaata | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tctctacaat aaagcaattt aaagcaattt aaagcaattt aagagaacaaa atcctgacca aaagctatag aggaacagaa acggacca aactggatat gcaaatcaaa aaaacaccgg tgcaggcaca tctgcccagc tattgcaaa | 4320 4380 4440 4500 4560 4620 4680 4740 4860 4920 5040 5160 5220 5280 5340 5400 5520 |
| gaggagaagt taataaaagc ctataagatc gaagtcctag gaagaagtca acaccactaa aaaaatcagt cagctacaaa gaaaactata caattctaag aaaagaacaa taacccaaac tagagaatcc agaacatact ccatatgcaa atggatgaaa atggatgaaa aggaacctcc ggcaaccaaa aaaggaaacc ct | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta agtatttcta taaaattagat gattggaaga tgcaatccct atttgtacag aactggaagc tacatggtac agaaacaaat ttggggaaa aataacaata ggcttaaatc caggacattg gcaaaaacag aacaagaat aacaaggaat | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt attaaatactg attaaaatac aaccacaaaa atcacattac tggcataaac ccatgcatct gataatctct ctagaactct taaaacctca gagtgggcaa acaaatgggg agaagagaca taataaccag | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaaaatgtc taatgacgtt gacccagaat ctgactcaa acagatgaga acagtgaact tcaataaatg gtctctcac aactttgcaa agacttcttg tcatatcaag acccacagaa tatatataag | agaagaaca ggaaaaactg tgtgattcaa aaggcatcca atatctggaa aggatacaaa tgaaaaagaa caaaaaagaa catactaccc cttcacagaa agccaaagct attatactac catggaccag cattittgac gtgctggagg atatacaaaa ctactaaaaa agtaattccc ttgggaggaata ggtaattccc | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tctctacaat aaagcaattt atagaagaaa atcctgacca aaagctatag aggaacagaa aaaggtgcca aactggatat gcaaatcaaa aaaacaccgg tgcaggcaca tctgcccagc tatttgcaaa tactctataa | 4320 4380 4440 4500 4560 4680 4740 4860 4920 4980 5100 5160 5220 5340 5460 5520 5580 |
| gaggagaagt taataaaagc ctataagatc gaagtcctag gaagaagtca acaccactaa aaaaatcagt cagctacaaa gaaaactata caattctaag aaaagaacaa taacccaaac tagagaatcc agaacatact ccatatgcaa atggatgaaa atggatgaaa aggaacctcc ggcaaccaaa aaaggaaacc ct | ggaagggaa cctatatgac tggaaaatga ctagagcaat aatatcctg aaaactatta agtatttcta taaaattaaa aaatattgat gattggaaga tgcaatccct aattgtacag aactggaagc tacatggtac agaaacaaa ttggggaaaa aataacaata ggcttaaatc caggacattg gcaaaacag actggaaacaaa | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt attaaatactg attaaaatac aaccacaaaa atcacattac tggcataaac ccatgcatct gataatctct ctagaactct taaaacctca gagtgggcaa acaaatgggg agaagagaca taataaccag | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaaaatgtc taatgacgtt gacccagaat ctgactcaa acagatgaga acagtgaact tcaataaatg gtctctcac aactttgcaa agacttcttg tcatatcaag acccacagaa tatatataag | agaagaaca ggaaaaactg tgtgattcaa aaggcatcca atatctggaa aggatacaaa tgaaaaagaa caaaaaagaa catactaccc cttcacagaa agccaaagct attatactac catggaccag cattittgac gtgctggagg atatacaaaa ctactaaaaa agtaattccc ttgggaggaata ggtaattccc | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tctctacaat aaagcaattt atagaagaaa atcctgacca aaagctatag aggaacagaa aaaggtgcca aactggatat gcaaatcaaa aaaacaccgg tgcaggcaca tctgcccagc tatttgcaaa tactctataa | 4320 4380 4440 4500 4560 4620 4680 4740 4860 4920 5040 5160 5220 5280 5340 5400 5520 |
| gaggagaagt taataaaagc catctaag gaagaagtca acaccactaa aaaaatcagt cagctacaaa gaaaactata caattctaag aaaagaacaa taacccaaaa tagagaatcc agaacatact ccatatgcaa atgagagatcc agaacatact ccatatgcaa aggagatcc ggcaaccaaa aggaaacct ggcaaccaaa ctactct ggcaaccaaa ctattcatct | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta agtatttcta taaaattaaa aaatattgat gattggaaga tgcaatccct atttgtacag aactggaagc tacatggtac agaaacaaat ttggggaaaa attaacaata ggcttaaatc caggacattc gcaaaacaag atcaacaaag atcaacaaag atcaacaaag | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccacac cagctaggaa aaaagaaatt ataaatact aaccacaaaa atcacattac tggcataaaa ccatgcatct gataatctct taaaactct taaaactct agatgggcaa acaatgggca acaaatggga agaagagaca ataataccag atttcaaaa | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaatagacgtt gacccagaat ctgacttcaa acagatgaga acaggtgaact tcaataaatg gttctcacc aagactttgt tcatacaa agacttcttg tcatatcaag acccacagaa accacagaa aacactagaa ataatataaa ataagcaaaa | agaagaaca ggaaaaacta ggaatcaa aaggcatcaa atatctggaa aggatacaaa aggtgaaaga caaaaaaga catactacco cttcacagaa agccaaagct attatactac catggaccag cattttgac gtgctggagg atatacaaaa ctactaacaaa gataattccc ttaaaaagct tgggggaata ggggggaata gggctcaaac gactctgggta | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tctctacaat aaagatattc aaagcaattt atagaagaaa atcctgacca aaagctatag aggaacagaa aaaggtgcca aactggatat gcaaatcaaa acaaccgg tgcaggcaca tctgcccag tattgcaaa tactctataa gacatttct | 4320 4380 4440 4500 4680 4740 4800 4920 4980 5040 5100 5220 5280 5340 5400 5520 5580 5540 |
| gaggagaagt taataaaagc ctctaagatc gaagtactag gaagaagtca acaccactaa aaaaaccagt cagctacaaa gaaaactata catgttcata acaattcaa caattctaag aaaagaacaa taacccaaaa caagaaacc tagagaatcc cgaaaacatact cgcaaccaaa aaggaaacc ggcaaccaa aaaggaaaca ctattcatc ggcaaccaa aaaggaaaca ctattcatc gaaaaacac caatatcac aaaagaaaca | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta agtatttcta taaaattaaa aaatattgat gattggaaga tgcaatccct atttgtacag aactggaagc tacatggaac tacatggaa atggaacaat ttggggaaaa aggaacaat ggcttaaata ggcttaaatc ggcaaaacag acaaggaa atcaacaag atcaacaag aacaaggaa taataagatg ataaaatgg | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt ataaatactg atcacaaaa atcacataac cagctagaac catgcatct gataatctt taaaactct cagatggca agagggcaa aatgggca agatggga agaagagaca tattaacaggca acataccaggca | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaacagtt gacccagaat ctgacttcaa acagatgagac tcaataaatg gtctctcac aactttgcac agactttcat gtctctcac agacttcata agacttcttg tcatacaag accacagaa atatataaaa atctgaaaat tctgaaaat | agaagaaca ggaaaaactg ggtgtgattcaa aaggcatcca atatctggaa aggatacaaa tgaaaaagaa catactaccc cttcacagaa agccaaagct attatactac catggaccag cattttgac gtgctggagg atatacaaaa ctactaaaagct tgggagaata gagtaattccc tgggagaata gagtccaacac | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tctctacaat aaagatattc aaagcaattt atagaagaaa atcctgacca aaagctatag aggaacagaa aaaggtgcca aactggatat gcaaatcaaa acaaccag tgcaggcaca tctgccagc tatttgcaaa tacttataa agacattttc | 4320 4380 4440 4560 4620 4620 4680 4740 4860 4980 5040 5100 5160 5220 5280 5340 5460 5520 5580 5580 5700 |
| gaggagaagt taataaaagc ctctaagatc gaagtactag gaagaagtca acaccactaa aaaaaccagt cagctacaaa gaaaactata catgttcata acaattcaa caattctaag aaaagaacaa taacccaaaa caagaaacc tagagaatcc cgaaaacatact cgcaaccaaa aaggaaacc ggcaaccaa aaaggaaaca ctattcatc ggcaaccaa aaaggaaaca ctattcatc gaaaaacac caatatcac aaaagaaaca | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta agtatttcta taaaattaaa aaatattgat gattggaaga tgcaatccct atttgtacag aactggaagc tacatggtac agaaacaaat ttggggaaaa attaacaata ggcttaaatc caggacattc gcaaaacaag atcaacaaag atcaacaaag atcaacaaag | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt ataaatactg atcacaaaa atcacataac cagctagaac catgcatct gataatctt taaaactct cagatggca agagggcaa aatgggca agatggga agaagagaca tattaacaggca acataccaggca | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaacagtt gacccagaat ctgacttcaa acagatgagac tcaataaatg gtctctcac aactttgcac agactttcat gccacagaa atatacaag ataagaaaa atataaaaa tctgaaaatg | agaagaaca ggaaaaactg ggtgtgattcaa aaggcatcca atatctggaa aggatacaaa tgaaaaagaa catactaccc cttcacagaa agccaaagct attatactac catggaccag cattttgac gtgctggagg atatacaaaa ctactaaaagct tgggagaata gagtaattccc tgggagaata gagtccaacac | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tctctacaat aaagatattc aaagcaattt atagaagaaa atcctgacca aaagctatag aggaacagaa aaaggtgcca aactggatat gcaaatcaaa acaaccag tgcaggcaca tctgccagc tatttgcaaa tacttataa agacattttc | 4320 4380 4440 4500 4680 4740 4800 4920 4980 5040 5100 5220 5280 5340 5400 5520 5580 5540 |
| gaggagaagt taataaaagc ctctaaggaccactaa aaaaaccactaa gaaaactata catgttcata acaaatcaaa caattctaag aaaagaacaa taaggaacaa taggagaacc ccatatgcaa aggaaacct ccatatgcaa aggagaacc ccatatcatc gacaacaaa caattcatc ccatatgcaa aggaaacct ccatatcatc cagaaaaccc cagaaaacac caaattcatct gaaaaacacc caaaataagt cagagaaat | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta agtatttcta taaaattaaa aaatattgat gattggaaga tgcaatccct aattgtacag aactggaagc tacatggtac agaaacaaa ttggggaaaa aataacaata ggcttaaatc caggacattg gcaaaacag atcaacaag atcaacaag atcaacaag ataataagct ataataagct cataaaatgg caaaacaag | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat cattccaaca cagctaggaa aaaagaaatt ataaaatactg aatcacacaaaa atcacataac tggcataaaa ccatgcatct ctagaactct taaaacccca gagggggca agaaggggca acaaatggga atttccaaaa actacctca cagatgggca caataaccag atttccaaa caacaggca ctactatgag | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaaaatgtc taatgacgtt taacgaaga ccagaat ctgacttcaa acagatgaga acagtgaga tcaataaatg gtctctcacc aactttgcaa agacttcttg tcatatcaag accacagaa tatataaag ataagcaaaa tctgaaaatg agatcatctc | agaagaaca ggaaaaactg tgtgattcaa aaggcatcca atatctggaa aggatacaaa tgaaaagaa caaaaaaaga catactaccc catggaccag cattittgac gtgctggagg atatacaaaa ctactaaaag agtaattcc ttgaaggataattcc ttggaggaata gagtattcgac gatctggag atatctagac agtattccaaca agctcaaac | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tctctacaat aaagcaattt aaagcaattt aaagcaattt aagaagaaa atcctgacca aaagctatag aggaacagaa aaggagcca aactggatat gcaaatcaaa aaaacacgg tgcaggcaca tctgcccagc tattgcaaa tacttataa gacattttc aactgatcat aactgatcat aactgatcat aactgatcat aactgatcat aactgatcat aaatggcttt | 4320 4380 4440 4560 4620 4680 4740 4860 4920 5040 5160 5220 5280 5340 5400 5520 5580 5640 5760 |
| gaggagaagt taataaaagc catcaagatc gaagtcctag gaagaagtca acaccactaa aaaaatcagt cagctacaaa gaaaactata acaagtcaa caattctaag aaaagaacca tagagaatcc agaacatact ccatatgcaa atggagagaatact cggcaaccaaa atggaaacaccaaa actatcatct ggcaaccaaa ctattcatct gaaaaacacc caaaataagtc caaaataagtc cagagaatg | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta agtatttcta taaaattaaa aaatattgat gattggaaga tgcaatccct atttgtacag aactggaagc tacatggtac agaaacaaa ttggggaaa attaacaata ggcttaaatc caggacattg gcaaaaacag aacaaggaat taataacatg gcaaaaacg aacaaggaat taataactg | ggggaaggga agaccgagggt caagggccca cagataagag tttgcagatg gagctgaaat tattccacac cagctaggaa aaaagaaatt attaaaatact attaccataac atgcacaaaa atcacattac tagaactc ttaaaacctc taaaacctc gagtgggcaa acaaatggga agaaagagaca taataaccag atttcaaaa ccatactaga actactatgag actactatgag actactatgag actactatgag acaaatggca | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagg gcaaacaatc ttaaccaaag gaagagggca ttaaacagtc taatgacgtt gacccagaat ctgacttcaa acagatgaga acagtgaac tcaataaatg gtctctcacc aactttgcaa agacttcttg tcatatcaag acacacgaa tatataaag atatataaag atagacaaaa tctgaaaatg tctgaaatg | agaagaaaca ggaaaaactg ggagatccaa aaggcatccaa aagtgaaaga caaaaaaaga aagtgaaagc cttcacagaa agccaaagctattatactac catggaccag cattttgac gtgctgagg atatacaaaagct tataaaaagct tagagacagctattaaaaagct tgggagaatatccc ttagagagagaattccc ttagagagagatattccc tagagagagattaggagaatagactccaagacagattgggta tgctcaaaaggattaaaaaggt | tattcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac aacaaaaaag tctctacaat aaagatattc atagaagaaa atcctgacca aaagctatag aggaacagaa aaagctatag cgaatcaacagaa aacggtgcca aactggatat gcaaatcaaca aaaacaccgg tgcaggcaca tctgcccagc tattgcaaa tacttataa gacatttcc cactgatcat aaagcattt | 4320 4380 4440 4500 4560 4680 4740 4860 4920 4980 5160 5220 5280 5340 5460 5520 5580 5640 5700 5760 5820 |
| gaggagaagt taataaaagc catcaagatc gaagtcctag gaagaagtca acaccactaa aaaaatcagt cagctacaaa gaaaactata acaagtcaa caattctaag aaaagaacca tagagaatcc agaacatact ccatatgcaa atggagagaatact cggcaaccaaa atggaaacaccaaa actatcatct ggcaaccaaa ctattcatct gaaaaacacc caaaataagtc caaaataagtc cagagaatg | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta agtatttcta taaaattaaa aaatattgat gattggaaga tgcaatccct aattgtacag aactggaagc tacatggtac agaaacaaa ttggggaaaa aataacaata ggcttaaatc caggacattg gcaaaacag atcaacaag atcaacaag atcaacaag ataataagct ataataagct cataaaatgg caaaacaag | ggggaaggga agaccgagggt caagggccca cagataagag tttgcagatg gagctgaaat tattccacac cagctaggaa aaaagaaatt attaaaatact attaccataac atgcacaaaa atcacattac tagaactc ttaaaacctc taaaacctc gagtgggcaa acaaatggga agaaagagaca taataaccag atttcaaaa ccatactaga actactatgag actactatgag actactatgag actactatgag acaaatggca | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagg gcaaacaatc ttaaccaaag gaagagggca ttaaacagtc taatgacgtt gacccagaat ctgacttcaa acagatgaga acagtgaac tcaataaatg gtctctcacc aactttgcaa agacttcttg tcatatcaag acacacgaa tatataaag atatataaag atagacaaaa tctgaaaatg tctgaaatg | agaagaaaca ggaaaaactg ggagatccaa aaggcatccaa aagtgaaaga caaaaaaaga aagtgaaagc cttcacagaa agccaaagctattatactac catggaccag cattttgac gtgctgagg atatacaaaagct tataaaaagct tagagacagctattaaaaagct tgggagaatatccc ttagagagagaattccc ttagagagagatattccc tagagagagattaggagaatagactccaagacagattgggta tgctcaaaaggattaaaaaggt | tattcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac aacaaaaaag tctctacaat aaagatattc atagaagaaa atcctgacca aaagctatag aggaacagaa aaagctatag cgaatcaacagaa aacggtgcca aactggatat gcaaatcaaca aaaacaccgg tgcaggcaca tctgcccagc tattgcaaa tacttataa gacatttcc cactgatcat aaagcattt | 4320 4380 4440 4560 4620 4680 4740 4860 4920 5040 5160 5220 5280 5340 5400 5520 5580 5640 5760 |
| gaggagaagt taataaaagc ctctaagatc gaagtcctag gaagaagtca acaccactaa aaaaatcagt cagctacaaa gaaaactata caattctaag aaaagaacaa taacccaaaa tagaagaatcc agaacatact cgatagaaaa atggagaatcc ggcaaccaaa atggagaaact ggcaaccaaa atggagaacc ggcaaccaaa aaaggaaacc ggcaaccaaa aaaggaaacc caattcatc gaaaaacac caattcatc gaaaaacac tagaagaaatg tattcaaaag acactgttgg | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaaactatta taaaattataa aaatattgat gattggaagc tacatggaagc tacatggaagc tacatggtaag aactaggaagc tacatggtaag aactaggaac ttgggagaaa aataacaata ggggaaat gcaaaacaag atcaacaag atcaacaag atcaacaag atcaacaag caaatcaaat | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt ataaatact aaccacaaaa atcacataac tggataatctc gataatctc taaaacctc agatgggcaa acaaatggga agaagagaca acaataacca taataacca taataacca taataacca actactataaga actactatagag agatgagca acaaatggga agaagagaca taataacca aatttcaaaa caacaggca ctactatgag acaaatgcca aattgcca aattgcca | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc ttaaccaaag gaagagggca ttaacaatgtc taatgacgtt gacccagaat ctgacttcaa acagatgaga acagtgaact tcaataaatg gtctctcacc agactttcta tcatacaag ataagacattcttg tcatacaaa acagatgaact tcaataaatg gtctctcacc agacttcttg tcatacaaa acagatgaact tcaatacaaa agacttcttg tcataccaaaa tctgaaaata atatataaaa tctgaaaata agaggatga actatggagat acattggagat actatggaga | agaagaaca ggaaaaacta ggaatcaa aaggcatcaa atatctggaa aggatacaaa aggtgaaaga caaaaaagaa aactactacca cttcacagaa agccaaagct attatactac catggaccag cattttggac gtgtggagg atatacaaaa agtaataccc ttgggagaata ggatcaaagct tgggagaata gagtcaaacac acacccagtta ggataaaagg acatttggata acccccagtta ggataaaagg acagtttgaa | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtaca accaaaaaag tctctacaat aaagatattc aaagcaattt atagaagaaa atcctgacca aaagctatag aggaacagaa aaaggtacca aactggatat gcaaatcaaa acaaccgg tgcaggcaca tctgcccagc tatttgcaaa tactctataa gacatttct cactgatcat aaatgcttt aaatgcttt aaactgctta aaatgcttt aaacccttgg agttcctcaa | 4320 4380 4440 4500 4620 4680 4740 4860 4920 4980 5040 5160 5220 5280 5340 5460 5520 5580 5640 5760 5760 5820 5880 |
| gaggagaagt taataaaagc ctctaagatc gaagtactag gaagaagtca acaccactaa aaaaatcagt cagctacaaa gaaaactata catgttcata acaattctaag aaaagaacaa taagcaaatcc agaaaccatact ccatatgcaa atggatgaaa ctattcatct ggcaaccaac ctattcatct ggaaaaccc aaaataagtc cagagaaatg tattcaaag aaacacc aaataagtc cagagaaatg tattcaaaag actgttgg acactgttgg aaaactaaaa | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaaactatta taaaattataa aaatattgat gattggaaga tgcaatccct atttgtacag aactggaagc tacatggtaa attgggaaaa ttggggaaaa attagggaaaa ttggggaaaa aataacaata ggcttaaatc gcaaaacag acaaggaat taaaaagct gaaacaaaa gacattg gcaaaaacag acaaggaat taataagct ataaaacg ataaaacga acaaggaat | ggggaaggga agaccgaggt caagggcca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt ataaatactg atcacaaaa atcacataac cggcataaaa ccatgcatct gataatctt taaaactct taaaactct taaaactct taaaactct aataactga agatggga agaagggaa agaagggaa caataaccag tattcaaaa caacaggca ctactatgag acaaatgcca ctactatgag acaaatgcca catacagcaa | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaacagtt gacccagaat ctgacttcaa acagatgaga ctgacttcaa acagatgaga tcaataaatg gtctctcacc aactttgcaa agacttcttg tcatatacag acccacagaa tatatataag atatatataag atatatataag atatacattc gtgaggattc gtgaggattc gagatctt gactattggaga acccattggaga acccattggaga acccattgct | agaagaaca ggaaaaact ggaatcca aaggcatcca atatctggaa aggatacaaa tgaaaagaa aagtgaaaga catactaccc cttcacagaa agccaaagct attatactac gtgctggagg atatacaaa ctactaaaag cattttgac gtgctggagg atatacaaag cttaaaag cttaaaag tctagaatcc ttaaaag agttcaaac ttaggacaag agttcaaac acccagtta ggataaaagg acagtttggaa gagtatatacaaaag tgctcaacac accccagtta ggataaaagg acagtttggaa aggtatataca | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tctctacaat taaagatattc aaagcaattt atagaagaaa atcctgacca aaagctatag aggaacagaa aaaggtgcca aactggatat gcaaatcaaa acaaccgg tgcaggcaca tctgccagc tatttgcaaa tactctataa agaatttct cactgatcat aaatgcttt aaacccttgg agttctcaa tcctagacat cactgatcat acctgatcat cactgatcat | 4320 4380 4440 4560 4620 4680 4740 4860 4980 5040 5100 5160 520 5280 5340 5400 5520 5580 5760 5760 5820 5880 5940 |
| gaggagaagt taataaaagc ctctaagatc gaagtactag gaagaagtca acaccactaa aaaaatcagt cagctacaaa gaaaactata catgttcata acaattctaag aaaagaacaa taagcaaatcc agaaaccatact ccatatgcaa atggatgaaa ctattcatct ggcaaccaac ctattcatct ggaaaaccc aaaataagtc cagagaaatg tattcaaag aaacacc aaataagtc cagagaaatg tattcaaaag actgttgg acactgttgg aaaactaaaa | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaaactatta taaaattataa aaatattgat gattggaagc tacatggaagc tacatggaagc tacatggtaag aactaggaagc tacatggtaag aactaggaac ttgggagaaa aataacaata ggggaaat gcaaaacaag atcaacaag atcaacaag atcaacaag atcaacaag caaatcaaat | ggggaaggga agaccgaggt caagggcca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt ataaatactg atcacaaaa atcacataac cggcataaaa ccatgcatct gataatctt taaaactct taaaactct taaaactct taaaactct aataactga agatggga agaagggaa agaagggaa caataaccag tattcaaaa caacaggca ctactatgag acaaatgcca ctactatgag acaaatgcca catacagcaa | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaacagtt gacccagaat ctgacttcaa acagatgaga ctgacttcaa acagatgaga tcaataaatg gtctctcacc aactttgcaa agacttcttg tcatatacag acccacagaa tatatataag atatatataag atatatataag atatacattc gtgaggattc gtgaggattc gagatctt gactattggaga acccattggaga acccattggaga acccattgct | agaagaaca ggaaaaact ggaatcca aaggcatcca atatctggaa aggatacaaa tgaaaagaa aagtgaaaga catactaccc cttcacagaa agccaaagct attatactac gtgctggagg atatacaaa ctactaaaag cattttgac gtgctggagg atatacaaag cttaaaag cttaaaag tctagaatcc ttaaaag agttcaaac ttaggacaag agttcaaac acccagtta ggataaaagg acagtttggaa gagtatatacaaaag tgctcaacac accccagtta ggataaaagg acagtttggaa aggtatataca | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tctctacaat taaagatattc aaagcaattt atagaagaaa atcctgacca aaagctatag aggaacagaa aaaggtgcca aactggatat gcaaatcaaa acaaccgg tgcaggcaca tctgccagc tatttgcaaa tactctataa agaatttct cactgatcat aaatgcttt aaacccttgg agttctcaa tcctagacat cactgatcat acctgatcat cactgatcat | 4320 4380 4440 4500 4620 4680 4740 4860 4920 4980 5040 5160 5220 5280 5340 5460 5520 5580 5640 5760 5760 5820 5880 |
| gaggagaagt taataaaagc ctctaaggaccataa aaaaaccactaa gaaaactata catgttcata gaaaactata caattctaag aaaagaacaa taacccaaac taggagaacca cagaacatacc ccatatgcaa aggaaacct ccatatgcaa aggagaacc cggcaaccaaa aaaggaaacc ctattcatct gaaaaaccc caaaatact cagagaaacg tattcatct gaaaacacc caaactagcaaa aaaggaaaca ctattcatct gacaacacaa aaaggaaaca ctattcatct gacaacacaa aaaggaaaca caaactact cagagaaatg tattcaaaag acactgttgg aaaactaaaa ggaatcagtg | ggaagggaa cctatatgac tggaaaatga ctagagcaat aatatcctg aaaactatta agtatttcta taaaattaaa aaatattgat gattggaaga tgcaatccct aattgtacag aactggaagc tacatggtac agaaacaaa ttggggaaaa aataacaata ggcttaaatc caggacattg gcaaaacag atcaacaag atcaacaag atcaacaag atcaacaag atcaacaag atcaacaag atcaacaag atcaacaag atcaacaag atcaacaag ataaaggcatta tagggaatgga ataaggcaatt | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt attaaaatac aaccacaaaa atcacatac tggcataaaa ccattac tggcataaaa ccatgcatct taaaacctc tgataatctct ttaaaacctca gagtgggca aacaatggaa atttccaaca actacagga actatacagga actatacagaa ccatcagaa ccatcagaa ccatcagaa ccatcagaa ccatcagaa ccatcagaa ccatcagcaa ctatctcca | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaaaatgtc taatgacgtt tagcccagaat ctgacttcaa acagatgaga acagtgaaca tcaataaatg gtctctcac aactttgcaa agacttcttg tcatatcaag acaagcaact tctgaaaatg agactccttg tcatatcaag atatataaag atatataaag atatatataag agatcactcc gtgaggatgt actatggagg tcccattgct tcccacattt | agaagaaca ggaaaaactg tgtgattcaa aaggcatcca atatctggaa aggatacaaa tgaaaaagaa caaaaaaaga catactaccc atgcaaagct attatactac catggaccag cattittgac gtgctggagg atataccaaa ctactaaaag agtaattcc ttggagaata gagctcaaac ttggtaatctggct tggagaata tgctcaacac gatctggta gatattcggt tggagaata ggatattcc acccagtta ggataaaaag aggatatacaac catggagaata aggataatacaac catggagaata | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tctctacaat aaagcaattt aaagcaattt aaagcaattt aagcaattta aagcaattag aggaacagaa acctgacca aaaggtgcca aactggatat gcaaatcaaa aaaacaccgg tgcaggcaca tctgcccagc tattgcaaa tacttataa gacatttct cactgatcat aaatggctt aaatggctt aaatggctt aaatggctt aaatgcttt aaacccttgg agttcctcaa tccaaaaaag tgttcataag | 4320 4380 4440 4560 4620 4680 4740 4860 4920 5040 5160 5220 5280 5340 5400 5760 5760 5820 5880 5940 6000 |
| gaggagaagt taataaaagc catcaagatc gaagtcctag gaagaagtca acaccactaa aaaaatcagt cagctacaaa gaaaactata catgttcata acaattctaag aaaagaacca tagagaatcc agaacatact ccatatgcaa atggaatcc ggcaaccaaa atggaatcc ggcaaccaaa ctattcatct ggcaaccaaa ctattcatct gaaaaacacc aaaatagtc cagagaatc cagagaatc ctattcatct gaaaaacacc aaaatagtc cagagaatg tattcaaaag acactgttgg aaaacctaatg acactgttgg aaaaccaggt | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta agtatttcta taaaattaaa aaatattgat gattggaaga tgcaatccct atttgtacag aactggaagc tacatggtac agaaacaaa ttggggaaaa attaacaattg gctaaaacag atcaacaaag aacaaggaat taataacatg caaacaaag aacaaggaat taataaactg ataaaactg ataaaactg ataaaacag acaacaaag aacaaggaat taataaactg ataaaacag caacaaaa aacaggcaata tgggaatgga ataaaacag tataaacaag | ggggaaggga agaccgagggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt attaaaatact attacactaca tggcataaaa ccatgcatct gataatctct ttaaaactcc gagtgggcaa acaaatgggca acaaatgggca taataaccag atttcaaaa ccatattac gattggcataaac ccatactagaa ctaataaccag atttcaaaa ccatattgag acaaatggca ctaattacac ctactatgag acaaatgcca actactgaa ccatcatgag ccaactaccac ccatcagtgtcc | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagg gcaaacaatc ttaaccaaag gaagagggca ttaaacagtc taatgacgtt gacccagaat ctgacttcaa acagatgaga acagtgaac tcaataaatg gtctctcac aactttgcaa agacttcttg tcatatacaag accacagaat atatataag accacagaat atatataag atatatatag atagcaaat ttgaaaatg tcgagatgat ctgagatgat ctcaccacgat tcatatcag accacagat tatataaag atatataaag | agaagaaaca ggaaaaacta ggaaaaacta gagatacaa agggatacaa agggaaaga aagtgaaaga caaaaaaga aagtgaaaga catactacc cttcacagaa agccaaagct attatactac catggaccag cattttgac gtgctggagg atatacaaaa ctactaaaag gtaattccc ttaaaaagct tgggagaat tgctcaaaa ggactcaaac gactcgggta tgctcaaaa gactcaaaag gactcaaac gactcgggta tgctcaaaa ggatattcc accccagtta ggataaaagg acagtttgaa aggtatatac actgcagcac gaatggaaaa ggataaaa | tattcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac aacaaaaaag tctctacaat aaagatattc atagaagaaa atcctgacca aaagctataag aggaacagaa aaagctatag caatttgcaaca acaggtgcca actggccca tctgcccagc tattgcaaat taatggacatttcataa gacatttctc cactgatcat gcaagcaaca tctgcccagc tattgcaaat taaatggcttt aaacccttgg agttcctcaa tccaaaaaaa tcctaaaa tccaaaaaa tcctaaaa tccaaaaaatgtg | 4320 4380 4440 4500 4680 4740 4860 4920 4980 5040 5160 5220 5280 5340 5400 5700 5760 5820 5880 5940 6000 6060 |
| gaggagaagt taataaaagc catcaagatc gaagtcctag gaagaagtca acaccactaa aaaaatcagt cagctacaaa gaaaactata catgttcata acaattctaag aaaagaacca tagagaatcc agaacatact ccatatgcaa atggaatcc ggcaaccaaa atggaatcc ggcaaccaaa ctattcatct ggcaaccaaa ctattcatct gaaaaacacc aaaatagtc cagagaatc cagagaatc ctattcatct gaaaaacacc aaaatagtc cagagaatg tattcaaaag acactgttgg aaaacctaatg acactgttgg aaaaccaggt | ggaagggaa cctatatgac tggaaaatga ctagagcaat aatatcctg aaaactatta agtatttcta taaaattaaa aaatattgat gattggaaga tgcaatccct aattgtacag aactggaagc tacatggtac agaaacaaa ttggggaaaa aataacaata ggcttaaatc caggacattg gcaaaacag atcaacaag atcaacaag atcaacaag atcaacaag atcaacaag atcaacaag atcaacaag atcaacaag atcaacaag atcaacaag ataaaggcatta tagggaatgga ataaggcaatt | ggggaaggga agaccgagggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt attaaaatact attacactaca tggcataaaa ccatgcatct gataatctct ttaaaactcc gagtgggcaa acaaatgggca acaaatgggca taataaccag atttcaaaa ccatattac gattggcataaac ccatactagaa ctaataaccag atttcaaaa ccatattgag acaaatggca ctaattacac ctactatgag acaaatgcca actactgaa ccatcatgag ccaactaccac ccatcagtgtcc | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagg gcaaacaatc ttaaccaaag gaagagggca ttaaacagtc taatgacgtt gacccagaat ctgacttcaa acagatgaga acagtgaac tcaataaatg gtctctcac aactttgcaa agacttcttg tcatatacaag accacagaat atatataag accacagaat atatataag atatatatag atagcaaat ttgaaaatg tcgagatgat ctgagatgat ctcaccacgat tcatatcag accacagat tatataaag atatataaag | agaagaaaca ggaaaaacta ggaaaaacta gagatacaa agggatacaa agggaaaga aagtgaaaga caaaaaaga aagtgaaaga catactacc cttcacagaa agccaaagct attatactac catggaccag cattttgac gtgctggagg atatacaaaa ctactaaaag gtaattccc ttaaaaagct tgggagaat tgctcaaaa ggactcaaac gactcgggta tgctcaaaa gactcaaaag gactcaaac gactcgggta tgctcaaaa ggatattcc accccagtta ggataaaagg acagtttgaa aggtatatac actgcagcac gaatggaaaa ggataaaa | tattcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac aacaaaaaag tctctacaat aaagatattc atagaagaaa atcctgacca aaagctataag aggaacagaa aaagctatag caatttgcaaca acaggtgcca actggccca tctgcccagc tattgcaaat taatggacatttcataa gacatttctc cactgatcat gcaagcaaca tctgcccagc tattgcaaat taaatggcttt aaacccttgg agttcctcaa tccaaaaaaa tcctaaaa tccaaaaaa tcctaaaa tccaaaaaatgtg | 4320 4380 4440 4560 4620 4680 4740 4860 4920 5040 5160 5220 5280 5340 5400 5760 5760 5820 5880 5940 6000 |
| gaggagaagt taataaaagc catctaagatc gaagtactag gaagaagtca acaccactaa aaaaatcagt cagctacaaa gaaaactata caattctaag aaaagaacaa taacccaaaa tagatgaaatcc agaacatact cgatatgaaa aggagaatcc ggcaaccaaa atggatgaaa caattcatc ggcaaccaaa ctattcatc gaaaaacac caattcatc gaaaaacac caattcatc gaaaaacac caattcatc gaaaaacac caattcatc gaaaaacac caaataagtc cagagaaatc tagagaaatc gaaaacac tagagaaatc gaaaacac caaaacacc aaaataagtc cagagaaatc gaaaacacac aaaccaaaa gaaacacacac | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaaactatta taaaattataa aaatattgat gattggaagc tacatggaagc tacatggaagc tacatggtac agaacaaat ttggggaaaa aataacaata ggcttaaatc caggacattg gcaaaacaag atcaacaag atcaacaag atcaacaag atcaacaag atcaacaag ataatagctt atacacaag ataaagcat tagggaattg gcaaatcaaa tgggaatgga ataaagctac tatcaacaag ataaagcaac acaaggaat | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt ataaatact aaccacaaaa atcacatacc tggataatctct ctagaactct taaaacctca gagtgggcaa acaaatggga agaagagaca tattaccaga atttcaaaa ccatggatcaa acaatagga acaaatggga agaaggaca acaatagga atttcaaaa caaccag attttcaaaa caaccag atttcaaa catactatgag acaaatggca ctactatgag acaaatgcca acatacagcaa ctatccc catacagcaa ctactccacc actaccag | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaatagattcaa cagatgaga acagtgaac tcaataaatg gtctctcacc agacttctag accattgaaaat tctgaaaatc tcatatacaag atatacaaaa atatataaaa agacttcttg tcatatcatg accacagaa tctgaaaat atatataaa agatcactc gtgagaat ctgaaaat ctgaaaat ctgaaaat ctgaaaat cataggaga tcccattgct tcccacatt tcccacatt accaacagac cataaaaaaa | agaagaaca ggaaaaacta ggaaaaacta ggatacaa aggatacaaa aggaaaga caaaaaaga catactacco cttcacagaa agccaaagct attatactac attggacaagct attatactac catggaccag cattttggac gtgctggagg atatacaaaa agtaataccactacaagaca actactaacac actcacagaa agtaattccc ttgggagaata gggtcaaac gactcaacac accccagtta ggataaagg acagttggaa aggtatatac acggacag aggtatatac acggacag aggtatatac acggacag | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tctctacaat aaagatattc aaagcaattt atagaagaaa atcctgacca aaagctatag aggaacagaa aaagctatag gcaaatcaaa acaccgg tctgccagc tattgcaaa tacttataa gcattttcaaag accttgacca tctgcccagc tattgcaaa tactctataa gacatttctc cactgatcat aaatggcttt aaacccttgg agttcctcaa tccaaaaaag tgttcataag tgttcataag tgttcatag tgttcatag | 4320 4380 4440 4500 4620 4680 4740 4860 4920 4980 5040 5160 5220 5280 5340 5460 5520 5580 5760 5760 5760 5880 5940 6060 6060 6120 |
| gaggagaagt taataaaagc ctctaagatc gaagcctag gaagaagtca acaccactaa acaaatcagt cagctacaaa gaaaactata catgttcata acaattctaag aaaagaacaa taacccaaaa caattctaag aaaacatact ccatatgcaa atggatgaaa ctgatgaaacc ggcaaccaaa aaaggaaacc ctattcatc ggcaaccaaa aaaggaaacc ctattcatc cagagaaatc tagaagaacc gaaacacc aaaataagtc cagagaaatg tattcaaaa ggaaacac catttcaaca gaaacacc aaaacacc aaaacacc aaaacacc aaaacacc aaaacacc aaaacacc aaaacacc aaaacacc aaacacc aaacacc aaaccc aaaacacc aaacacc aacacc aacacc aaaacacc aacacc | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaactatta taatattcta taaaattataa aaatattgat gattggaaga tgcaatccct atttgtacag aactggaagc tacatggaag tacatggaaa ttggggaaa attagggaaa ttggggaaaa ataacaata ggcttaaatc caggaacatg gcaaatcaaa aacaaggaat taataagctg atacaaagg atcaaaagg atcaaaagg atcaaaagg atcaaaagg ataaagcaat tgggaaatga acaggcaat tagggaatgg aaaagcaac acaggcaat | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt ataaatactg atcacaaaa atcacataac tggcataaaa ccatgcatct gataatctt taaaactca gagtgggaa agaaggca taataaccag agatggga agaagagaca taataaccag atttcaaaa caacaggca ctactatgag acaatgcac ctactatgag ctactatgcac ctacaacgca ctactaccc catacagca ctactccac catacgcac ctactacgcac | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt trggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaacaatgtc taatgacgtt gacccagaat ctgacttcaa acagatgaga acagtgaga ttcaataaatg gtctctcac aactttgcaa agactttgcaa agactttgcaa atatataaag atcagaaaat tcgaaaatg tcgaaaatg tcgagaatg tcgagaatg tcgagaatg atatgagaa atctgaaaatg agatcatctc gtgaggatg tcccattgct tcccacattt accaacagaa aagaggaaaaa | agaagaaca ggaaaaactg tgtgattcaa aaggcatcca atatctggaa aggatacaaa tgaaaaga catactaccc cttcacagaa agccaaagct attatactac catggaccag cattttgac gtgctggagg atatacaaaa ctactaaaac tactaaaagct tgggagaata gagtcaaagct tgggagaata gagtcaaagct aggatacac acccagtta ggataaaagg acagtttggaa aggataaaag acgattggaa aggataaaaa aggataaaaa aggataaaaa aggataaaa aggataac cacgaaggaaa aaggaaaa aatggaaaa aatggaaaaa aatgagaacc agccaggcac | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tctctacaat aaagatattc aaagcaattt atagaagaaa atcctgacca aaagctatag aggaacagaa aaaggtgcca aactggatat gcaaatcaaa acaaccgg tgcaggcaca tctctacaa tactctataa acactttta aaatgcttt aaactcttata aatgcttt cactgatcat aaatgcttt aaatgcttt aaacccttgg tgtcagtcat tctcaaaaaag tgttcatag tgtcaaaacagcgg tgtcagatctcaa | 4320 4380 4440 4560 4620 4680 4740 4860 4980 5040 5100 5120 5280 5340 5400 5520 5580 5760 5760 5820 5820 6060 6060 6120 6180 |
| gaggagaagt taataaaagc ctctaagatc gaagcctag gaagaagtca acaccactaa acaaatcagt cagctacaaa gaaaactata catgttcata acaattctaag aaaagaacaa taacccaaaa caattctaag aaaacatact ccatatgcaa atggatgaaa ctgatgaaacc ggcaaccaaa aaaggaaacc ctattcatc ggcaaccaaa aaaggaaacc ctattcatc cagagaaatc tagaagaacc gaaacacc aaaataagtc cagagaaatg tattcaaaa ggaaacac catttcaaca gaaacacc aaaacacc aaaacacc aaaacacc aaaacacc aaaacacc aaaacacc aaaacacc aaaacacc aaacacc aaacacc aaaccc aaaacacc aaacacc aacacc aacacc aaaacacc aacacc | ggaagggaa cctatatgac tggaaaatga ctagagcaat aattatcctg aaaaactatta taaaattataa aaatattgat gattggaagc tacatggaagc tacatggaagc tacatggtac agaacaaat ttggggaaaa aataacaata ggcttaaatc caggacattg gcaaaacaag atcaacaag atcaacaag atcaacaag atcaacaag atcaacaag ataatagctt atacacaag ataaagcat tagggaattg gcaaatcaaa tgggaatgga ataaagctac tatcaacaag ataaagcaac acaaggaat | ggggaaggga agaccgaggt caagggccca cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt ataaatactg atcacaaaa atcacataac tggcataaaa ccatgcatct gataatctt taaaactca gagtgggaa agaaggca taataaccag agatggga agaagagaca taataaccag atttcaaaa caacaggca ctactatgag acaatgcac ctactatgag ctactatgcac ctacaacgca ctactaccc catacagca ctactccac catacgcac ctactacgcac | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt trggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaacaatgtc taatgacgtt gacccagaat ctgacttcaa acagatgaga acagtgaga ttcaataaatg gtctctcac aactttgcaa agactttgcaa agactttgcaa atatataaag atcagaaaat tcgaaaatg tcgaaaatg tcgagaatg tcgagaatg tcgagaatg atatgagaa atctgaaaatg agatcatctc gtgaggatg tcccattgct tcccacattt accaacagaa aagaggaaaaa | agaagaaca ggaaaaactg tgtgattcaa aaggcatcca atatctggaa aggatacaaa tgaaaaga catactaccc cttcacagaa agccaaagct attatactac catggaccag cattttgac gtgctggagg atatacaaaa ctactaaaac tactaaaagct tgggagaata gagtcaaagct tgggagaata gagtcaaagct aggatacac acccagtta ggataaaagg acagtttggaa aggataaaag acgattggaa aggataaaaa aggataaaaa aggataaaaa aggataaaa aggataac cacgaaggaaa aaggaaaa aatggaaaa aatggaaaaa aatgagaacc agccaggcac | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaaaaaag tctctacaat aaagatattc aaagcaattt atagaagaaa atcctgacca aaagctatag aggaacagaa aaaggtgcca aactggatat gcaaatcaaa acaaccgg tgcaggcaca tctctacaa tactctataa acactttta aaatgcttt aaactcttata aatgcttt cactgatcat aaatgcttt aaatgcttt aaacccttgg tgtcagtcat tctcaaaaaag tgttcatag tgtcaaaacagcgg tgtcagatctcaa | 4320 4380 4440 4500 4620 4680 4740 4860 4920 4980 5040 5160 5220 5280 5340 5460 5520 5580 5760 5760 5760 5880 5940 6060 6060 6120 |
| gaggagaagt taataaaagc ctctaaggaccactaa aaaaactata cagctacaaa gaaaactata caattctaag aaaaactata caattctaag aaaacaac taggagaaca taggagaacc ccatatgcaa atggatgaaa agaaactct ccatatgcaa atggatgaaa agaaacctct ggcaaccaaa ctattcatct ggaaaacac ctattcatct gaaaaacac caaaattaag cactgttga aacggaaac gaaccaggaaacg catatcaaag acactgtga aacactgtga aacactgtga aacacaggaaacg ccaggaaatg tattcaaaag acactgttga aacacgatgg acaccaggtt gtgcacata aacagcatgg acttttcatg | ggaagggaa cctatatgac tggaaaatga ctagagcaat aatatcctg aaaactatta aagtattcta taaaattaaa aaatattgat gattggaagc tacatggtac aactggaagc tacatggtac agaaacaaa ttggggaaaa aataacaat ggctaaaacag gcaaacaag acaaggaat taaaaggaat taaaaaggaat taaaaaggaat taaaaggaat taaaaggaat taaaaggaat taaaaggaat taaaaggaat taaaaggaat taaaaggaat taaaaggaat taaaaggaat taaaaggaat taaaaggaat taaaaggaat taaaaggaat | ggggaaggga agaccgaggt caagggccc cagataagag tttgcagatg gagctgaaat tattccaaca cagctaggaa aaaagaaatt ataaaataca atcacacaaaa atcacatac tggcataaaa ccatgcatct taaaacctc gataactct taaaacctc agagggggca aataaccag atttcaaaa accaatac tctagaactct taaaaccca gagggggca caataaccag atttcaaaa caatagcac ctatagag acaaatgcac ctatagag acaaatgcca catacagcaa ctactccac catcagcaa ctactccac catcagcaa ctactccac catcagcaa ctacgcac ctcagtgcc actagggcac actacgcac ctcagtagtt cttgtgggag | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatctt ttggtacagc gcaaacaatc ttaaccaaag gaagagggca ttaaaatgacgtt gacccagaat ctgacttcaa acagatgaga ccaagtgaga tcaataaatg gtctctcac aactttgcaa agacttcttg tcatatcaag acactcttg acccacagaa tatatataag ataagcaaaa tctgaaaatg agatcatctc gtgaggagga tactaggagga tactaggagaga tcccattgct tcccacattt atcaacaaaa aagtgaaata cataaaaaaa aagatgaaata | agaagaaca ggaaaaaca gagaacaca atactggaa aggatacaaa tgaaaagaa aaagtgaaaga catactaccc attactaccc attatactaccc attatactaccc attatactaccc attatactaccc attatactaccc attatactaccc tggagagaaca agcatatttagac gagtaatcacaaagct tgggagaata gagtatatcc acccagtta ggataaaagg acattggga acattggga acattggga acattggga acattgggataatacaac gattggagaata tgctcaaca accccagtta ggataaaagg acagttgga aaggtatatac acgattggaaaa aggtatatac acagagaca aatggaaca aatggaaca aacaattgac aacaattgac | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac accaacaaaag tctctacaat aaagatattc aaagcaattt atagaagaaa atcctgacca aaagctatag aggaacagaa aaaggtgcca aactggatat gcaatcaaa aaaacaccgg tgcaggcaca tctgcccagc tattgcaaa tacttataa gacattttc aaactctataa gacatttct cactgatcat aaatgcttt aaacccttgg agttcctcaa tccacaaaaaag tgtcatagc agaaaatgtg tgtagttg agaaaagcaa atagaaatag | 4320 4380 4440 4560 4620 4680 4740 4860 4920 5040 5160 5220 5340 5400 5400 5760 5820 5880 5760 5820 5880 5940 6060 6120 6180 6240 |
| gaggagaagt taataaaagc cactatagaaaccacactaa aaaaatcaat cagttacaa gaaaactata catgttcata acaactcaaa gaaaaccaa taacccaacaa taacccaacac tagagaatcc agaacatact ccatatgcaa aaggaaacct cggcaaccaaa aaggaaaccc ggcaaccaaa aaggaaacct ccatatgcaa aaggaaacct ccatatgcaa aaggaaacct ggcaaccaaa aaggaaacc caatagcaa aggaatcacc gaaaaacacc aaaatagtc cagagaaatg tattcaaaag acactgttgg aaaaactaagt agcaaggtt gtgcacaaggt ggcaaaggt agccaaggt agccaaggt agcaaggaatg agcacatac | ggaagggaa cctatatgac ctagagcaat aattatcctg aaaactatta agtatttcta taaaattaaa aaatattgat gattggaaga tgcaatccct atttgtacag aactggaaga tcaggaacaat ttggggaaaa aataacaata ggcttaaatc caggacattg gcaaaacaag aacaaggaat taacaaaag aacaaggaat taacaaaag caacaaaa aataacaat ggcaatcaaaa aacaggaat taacaaaag caacaaaa gaacaaggaat taacaaaag caacaaaa gaacaaggaat taacaaaag caacaaaa ggaatgaacaa taacaaaag taacaaaag caacaaaa tgggaatgga ataaaagctac tatcaacaag tgggaacagg tggggacacgg ggggcactgg | ggggaaggga agaccgagggt caagggccca cagataagag tttgcagatg gagctgaaat tattccacac cagctaggaa aaaagaaatt aatacacttag attaaaatacct gagcataaaa ccatgcatct gataatctct ctagaactct taaaacctca gagtgggcaa acaaatggga agaagagaca taataaccag atticaaaa cctactatgag acaaatggca acaatggca ctaataaccag atticaaaa cctactatgag acaaatgcca actactagag ccatacagcac ctactacgcag ctactacgcag ctactacgcag ctacgtactc actacgcag ctcagtacg gggggggggg | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatcat ttggtacagg gcaaacaatc ttaaccaaag gaagagggca ttaaacagat ctaatgacgtt gacccagaat ctgacttcaa acagatgaga acagatgaga acagatgaga acagatgaact tcaatacaag gaccacacactttgcaa agactcttg tcatatacaag accacagaat atatataag accacagaat tctgaaaatg agatcatctc gtgaggatgt actatggaga tccattgct tcccacattt atcaacagac cataaaaaaa aagtgaaata acagagtgaac | agaagaaca ggaaaaacta ggaaaaaca aaggcatcaa agggatacaaa tgaaaaagaa aagtgaaaga caaaaaagac catctaccac catctaccacaa agccaaagct attatactac catggacag catttttgac gtgctggagg atatacaaaa ctactaaaagct tagactaacacac ctacagaaagtaattccc ttagagagaata tgctcaacac gactcaggagaat ggataaaagg acagtttgaa aggtatatac accccagtta ggataaaagg acagtttgaa aggtatatac acatgagaaca gaatgagaaca gaatgagaaca gaatgagaaca gacaggcacagccag aacaattgac tagagtcaac tagagtcaac | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac aacaaaaaag tctctacaat aaagatattc aaagcaattt atagaagaaa atcctgacca aaagctatag aggaacagaa aaagctgatat gcaaatcaaa aaaacacaga tgcaggcaca tctgcccagc tattgcaaa tcattgcaaa tacttataa gacatttct cactgatcat aaaggctatag tgcaggcaca tctgcccagc tatttgcaaa tactctataa gacatttctc cactgatcat aaatggcatt gcaggcaca tctgcccagc tattgcaaa tccataaaa gacattctc cactgatcat aaatggctt aaacccttgg ggtcctcaa tccaaaaaaag tgtcaaagacaaa tgcaaaagacaa atagaaaatgtg tgtcagttgc agaaaagacaa aatagaaatag aataaatttat | 4320 4380 4440 4500 4680 4740 4860 4920 4980 5040 5160 5220 5280 5340 5460 5700 5760 5820 5880 5940 6060 6120 6180 6240 6300 |
| gaggagaagt taataaaagc cactatagaaaccacactaa aaaaatcaat cagttacaa gaaaactata catgttcata acaactcaaa gaaaaccaa taacccaacaa taacccaacac tagagaatcc agaacatact ccatatgcaa aaggaaacct cggcaaccaaa aaggaaaccc ggcaaccaaa aaggaaacct ccatatgcaa aaggaaacct ccatatgcaa aaggaaacct ggcaaccaaa aaggaaacc caatagcaa aggaatcacc gaaaaacacc aaaatagtc cagagaaatg tattcaaaag acactgttgg aaaaactaagt agcaaggtt gtgcacaaggt ggcaaaggt agccaaggt agccaaggt agcaaggaatg agcacatac | ggaagggaa cctatatgac tggaaaatga ctagagcaat aatatcctg aaaactatta aagtattcta taaaattaaa aaatattgat gattggaagc tacatggtac aactggaagc tacatggtac agaaacaaa ttggggaaaa aataacaat ggctaaaacag gcaaacaag acaaggaat taaaaggaat taaaaaggaat taaaaaggaat taaaaggaat taaaaggaat taaaaggaat taaaaggaat taaaaggaat taaaaggaat taaaaggaat taaaaggaat taaaaggaat taaaaggaat taaaaggaat taaaaggaat taaaaggaat | ggggaaggga agaccgagggt caagggccca cagataagag tttgcagatg gagctgaaat tattccacac cagctaggaa aaaagaaatt aatacacttag attaaaatacct gagcataaaa ccatgcatct gataatctct ctagaactct taaaacctca gagtgggcaa acaaatggga agaagagaca taataaccag atticaaaa cctactatgag acaaatggca acaatggca ctaataaccag atticaaaa cctactatgag acaaatgcca actactagag ccatacagcac ctactacgcag ctactacgcag ctactacgcag ctacgtactc actacgcag ctcagtacg gggggggggg | aagaggaaga agtattatga ctttcaccac aaagaaataa atatgatcat ttggtacagg gcaaacaatc ttaaccaaag gaagagggca ttaaacagat ctaatgacgtt gacccagaat ctgacttcaa acagatgaga acagatgaga acagatgaga acagatgaact tcaatacaag gaccacacactttgcaa agactcttg tcatatacaag accacagaat atatataag accacagaat tctgaaaatg agatcatctc gtgaggatgt actatggaga tccattgct tcccacattt atcaacagac cataaaaaaa aagtgaaata acagagtgaac | agaagaaca ggaaaaacta ggaaaaaca aaggcatcaa agggatacaaa tgaaaaagaa aagtgaaaga caaaaaagac catctaccac catctaccacaa agccaaagct attatactac catggacag catttttgac gtgctggagg atatacaaaa ctactaaaagct tagactaacacac ctacagaaagtaattccc ttagagagaata tgctcaacac gactcaggagaat ggataaaagg acagtttgaa aggtatatac accccagtta ggataaaagg acagtttgaa aggtatatac acatgagaaca gaatgagaaca gaatgagaaca gaatgagaaca gacaggcacagccag aacaattgac tagagtcaac tagagtcaac | tatttcaaca aaagcctttc catagtacta aactggaaag aagacttaag atcaatgtac aacaaaaaag tctctacaat aaagatattc aaagcaattt atagaagaaa atcctgacca aaagctatag aggaacagaa aaagctgatat gcaaatcaaa aaaacacaga tgcaggcaca tctgcccagc tattgcaaa tcattgcaaa tacttataa gacatttct cactgatcat aaaggctatag tgcaggcaca tctgcccagc tatttgcaaa tactctataa gacatttctc cactgatcat aaatggcatt gcaggcaca tctgcccagc tattgcaaa tccataaaa gacattctc cactgatcat aaatggctt aaacccttgg ggtcctcaa tccaaaaaaag tgtcaaagacaaa tgcaaaagacaa atagaaaatgtg tgtcagttgc agaaaagacaa aatagaaatag aataaatttat | 4320 4380 4440 4560 4620 4680 4740 4860 4920 5040 5160 5220 5340 5400 5400 5760 5820 5880 5760 5820 5880 5600 6060 6120 6180 6240 |

| | | | | ++-+ | | 6420 |
|---|--|---|---|--|---|--|
| | aggtgacaga | | | | | 6420 |
| ctgtatcaaa | atatctcatg | tatgctatag | atataaaccc | tactatatta | aaaattaaaa | 6480 |
| ttttaatggc | caggcacggt | ggctcatgtc | cataatccca | gcactttggg | aggccgaggc | 6540 |
| | ctgaggtcag | | | | | 6600 |
| | atacaaaaat | | | | | 6660 |
| | | | | | | |
| | agacaggaga | | | | | 6720 |
| atcatgccac | tgcactgcag | cctgggtgac | agagcaagac | tccatctcaa | aacaaaaaca | 6780 |
| aaaaaaaaaa | gattaaaatt | otaattttta | totaccotat | aaatatatac | tctactatat | 6840 |
| ****** | | ++-+ | | cacttaatet | 222222222 | 6900 |
| lagaayilaa | aaattaaaac | dattataaaa | ggtaattaat | Caccidates | aadataayaa | |
| caatgtatgt | ggggtttcta | gcttctgaag | aagtaaaagt | tatggccacg | atggcagaaa | 6960 |
| totoaggagg | gaacagtgga | agttactgtt | gttagacgct | catactctct | gtaagtgact | 7020 |
| | caaagacagg | | | | | 7080 |
| | | | | | | 7140 |
| | atggtgaaag | | | | | |
| aatcgagctg | cagaattggc | acgtctgatc | acaccgtcct | ctcattcacg | gtgctttttt | 7200 |
| tcttatatac | ttggagattt | tcgattgtgt | attcatattt | ggttaaactt | aatctgtatg | 7260 |
| | cgaaaaatgg | | | | | 7320 |
| | | | | | | 7380 |
| | ctgtggacct | | | | | |
| gtgcaaggca | gaggcctgat | gacccgagga | caggaaagct | cggatgggaa | ggggcgatga | 7440 |
| gaagcctgcc | tcgttggtga | gcagcgcatg | aaqtqccctt | atttacgctt | tgcaaagatt | 7500 |
| | ccatctggaa | | | | | 7560 |
| | | | | | | 7620 |
| | ggccagcagc | | | | | |
| aaggcacctc | gaagtatggc | ttaaatcttt | ttttcacctg | aagcagtgac | caaggtgtat | 7680 |
| | gcttgagtta | | | | | 7740 |
| | accagacgcc | | | | | 7800 |
| LCaayyyaaa | accagacycc | cyctctgtgg | ccacccaccc | | | |
| | tttctgatcg | | | | | 7860 |
| gaggaccctc | ttgcaaaggg | ctccacagac | ccccgccctg | gagagaggag | tctgagcctg | 7920 |
| octtaataac | aaactgggat | ataactaaaa | acadacaded | acoocoooat | tcaaagactt | 7980 |
| anttecates | atacattesa | 5-555555 cc+++cc1c1 | ********** | tttanatttt | atcttaatat | 8040 |
| | gtaaattcaa | | | | | |
| | ttcatcaaat | | | | | 8100 |
| actgagctat | gtttgccaag | gtccaaggac | ttaataacca | tgttcagagg | gatttttcgc | 8160 |
| | ttttattggt | | | | | 8220 |
| | | | | | | 8280 |
| | ggcggcaggg | | | | | |
| | gacagcagga | | | | | 8340 |
| gcgaaggcgg | ccacgctgcg | tgtgactcag | gaccccatac | cggcttcctg | ggcccaccca | 8400 |
| | ggaagtcacg | | | | | 8460 |
| | | | | | | 8520 |
| ctgcttccct | gggtgggtca | aggytaatya | agtygtgtgt | ayyaaatyyt | catytaaatt | |
| acacgactct | gctgatgggg | accgttcctt | ccatcattat | tcatcttcac | ccccaaggac | 8580 |
| tgaatgattc | cagcaacttc | ttcgqqtgtg | acaagccatg | acaacactca | gtacaaacac | 8640 |
| cactetttta | ctaggcccac | agaggaggg | ccacacccct | gatatattaa | gagtccagga | 8700 |
| | tastttass | agageaegge | castaccocc | 300000000 | contetatte | 8760 |
| | tgctttcagc | | | | | |
| | gtagaccctg | | | | | 8820 |
| ccaaaaacac | catctgccct | ggagactcag | cctggggtgc | cacactgagg | ccagccctgt | 8880 |
| | tccgcctcca | | | | | 8940 |
| | | | | | | 9000 |
| | gcgctactgt | | | | | |
| | tcctcacatg | | | | | 9060 |
| ggaggagatt | ctgcgcctcc | cagactggct | cctctgagcc | tgaacctggc | tcgtggcccc | 9120 |
| | tcctggcgtc | | | | | 9180 |
| cyatycayyt | ccctggcgtc | cygeryeary | tattatat | ctatactcat | ttccacatcc | 9240 |
| cctgtcatct | gccggggcct | gccggrgrgr | icittigiti | cigigeteet | ceccacycec | |
| | gtctctgtcc | | | | | 9300 |
| tggtgggcca | gggcgctctt | gggaaatgca | acatttgggt | gtgaaagtag | gagtgcctgt | 9360 |
| cctcacctag | gtccacgggc | acaggeetgg | ggatggagcc | cccaccagaa | acccoccctt | 9420 |
| | cacttttctg | | | | | 9480 |
| | | | | | | |
| | cctcttccca | | | | | 9540 |
| ccctgggaat | tcacgtgact | acgcacatca | . tgtacacact | cccgtccacg | accgaccccc | 9600 |
| octottttat | tttaatagct | acaaaqcaqq | gaaatccctg | ctaaaatgtc | ctttaacaaa | 9660 |
| ctaattaaac | aaacgggtcc | atconcacno | tanacantto | ctcacagiga | anannaacat | 9720 |
| ceggetaaac | anacygyccc | tetes | - cygucugica | tacatagaga | ctaccaccta | 9780 |
| gccgtttata | aagcctgcag | ycarcicaay | ggaarracyc | Lyayttaaaa | Cigicaccic | |
| catgggatac | gtacgcaaca | tgctcaaaaa | gaaagaattt | caccccatgg | caggggagtg | 9840 |
| gttgggagat | taaggacggt | gggggcagca | gctgggggct | actgcacgca | ccttttacta | 9900 |
| | cctggttctg | | | | | 9960 |
| | ggaacccgga | | | | | 10020 |
| | | | | | | |
| | agagatgccc | | | | | 10080 |
| gccccagggc | ctttgcaggt | gtgatctccg | tgaggaccct | gaggtctggg | atccttcggg | 1 01 40 |
| actacctoca | gocccgaaaa | otaatccago | gattctagga | agaggcgggc | aggagggtca | 10200 |
| | gcctcaggac | | | | | 10260 |
| | | | | | | 10320 |
| cctcgagccc | aggcctgcaa | gegeetecag | aagciggaaa | aaycggggaa | gggaccctcc | |
| acggagcctg | cagcaggaag | gcacggctgg | cccttagccc | accagggccc | atcgtggacc | 10380 |
| tccggcctcc | | | cactacactt | ctaccatasa | gtatataggg | 10440 |
| | gtgccatagg | agggcactcg | LULLUCCELL | LLagialyaa | | |
| | gtgccatagg | | | | | 10500 |
| | gtgccatagg gcaacaggaa | acccatgcac | : tgtgaatcta | ggattatttc | aaaacaaagg | 10500 |
| | gtgccatagg gcaacaggaa catccaagga | acccatgcac cagggctgaa | tgtgaatcta gtgcctccgg | ggattatttc gcaagggcag | aaaacaaagg ggcaggcacg | 10560 |
| | gtgccatagg gcaacaggaa catccaagga tttagctatt | acccatgcac cagggctgaa ttattttatt | tgtgaatcta gtgcctccgg tacttacttt | ggattattto gcaagggcag ctgagacaga | aaaacaaagg ggcaggcacg gttatgctct | 10560 10620 |
| | gtgccatagg gcaacaggaa catccaagga tttagctatt | acccatgcac cagggctgaa ttattttatt | tgtgaatcta gtgcctccgg tacttacttt | ggattattto gcaagggcag ctgagacaga | aaaacaaagg ggcaggcacg | 10560 |
| tgttgcccag | gtgccatagg gcaacaggaa catccaagga tttagctatt gctggagtgc | acccatgcac cagggctgaa ttattttatt agcggcatga | tgtgaatcta gtgcctccgg tacttacttt tcttggctca | ggattattto gcaagggcag ctgagacaga ctgcaaccto | aaaacaaagg ggcaggcacg gttatgctct cgtctcctgg | 10560 10620 |
| tgttgcccag gttcaagcaa | gtgccatagg gcaacaggaa catccaagga tttagctatt gctggagtgc ttctcgtgcc | acccatgcac cagggctgaa ttattttatt agcggcatga tcagcctccc | tgtgaatcta gtgcctccgg tacttacttt tcttggctca aggtagctgg | ggattatto gcaagggcag ctgagacaga ctgcaaccto gatttcaggo | aaaacaaagg ggcaggcacg gttatgctct cgtctcctgg gtgcaccacc | 10560 10620 10680 10740 |
| tgttgcccag gttcaagcaa acacccggct | gtgccatagg gcaacaggaa catccaagga tttagctatt gctggagtgc ttctcgtgcc aattttgtat | acccatgcac cagggctgaa ttattttatt agcggcatga tcagcctccc ttttagtaga | tgtgaatcta gtgcctccgg tacttacttt tcttggctca agtagctgg gatgggcttt | ggattattto gcaagggcag ctgagacaga ctgcaaccto gatttcaggo caccatgttg | aaaacaaagg ggcaggcacg gttatgctct cgtctcctgg gtgcaccacc gtcaggctga | 10560 10620 10680 10740 10800 |
| tgttgcccag gttcaagcaa acacccggct tctcaaaatc | gtgccatagg gcaacaggaa catccaagga tttagctatt gctggagtgc ttctcgtgcc aattttgtat | acccatgcac cagggctgaa ttattttatt agcggcatga tcagcctccc ttttagtaga gtgatccgcc | tgtgaatcta gtgcctccgg tacttacttt tcttggctca aggtagctgg gatgggcttt cacctcagcc | ggattattto gcaagggcag ctgagacaga ctgcaaccto gatttcaggo caccatgttg tcccaaagtg | aaaacaaagg ggcaggcacg gttatgctct cgtctcctgg gtgcaccacc gtcaggctga ctgggattac | 10560 10620 10680 10740 10800 10860 |
| tgttgcccag gttcaagcaa acacccggct tctcaaaatc | gtgccatagg gcaacaggaa catccaagga tttagctatt gctggagtgc ttctcgtgcc aattttgtat | acccatgcac cagggctgaa ttattttatt agcggcatga tcagcctccc ttttagtaga gtgatccgcc | tgtgaatcta gtgcctccgg tacttacttt tcttggctca aggtagctgg gatgggcttt cacctcagcc | ggattattto gcaagggcag ctgagacaga ctgcaaccto gatttcaggo caccatgttg tcccaaagtg | aaaacaaagg ggcaggcacg gttatgctct cgtctcctgg gtgcaccacc gtcaggctga | 10560 10620 10680 10740 10800 |
| tgttgcccag gttcaagcaa acacccggct tctcaaaatc aggcatgagc | gtgccatagg gcaacaggaa catccaagga tttagctatt gctggagtgc ttctcgtgcc aattttgtat ctgacctcag | acccatgcac cagggctgaa ttattttatt agcggcatga tcagcctccc ttttagtaga gtgatccgcc ggcctattta | tgtgaatcta gtgcctccgg tacttacttt tcttggctca aggtagctgg gatgggcttt cacctcagcc | ggattattto gcaagggcag ctgagacaga ctgcaaccto gatttcaggo caccatgttg tcccaaagtg | aaaacaaagg ggcaggcacg gttatgctct cgtctcctgg gtgcaccacc gtcaggctga ctgggattac ggctcaagtc | 10560 10620 10680 10740 10800 10860 |
| tgttgcccag gttcaagcaa acacccggct tctcaaaatc aggcatgagc acacccactg | gtgccatagg gcaacagga catccaagga tttagctatt gctggagtgc ttctcgtgcc aattttgtat ctgacctcag cactgcacct | acccatgcac cagggctgaa ttattttatt agcggcatga tcagcctcc ttttagtaga gtgatccgcc ggcctattta catggagttc | tgtgaatcta gtgcctccgg tacttacttt tcttggctca aagtagctgt gatgggcttt cacctcagcc accatttaa | ggattattto gcaagggcag ctgagacaga ctgcaaccto gatttcaggg ccaccatgttg tcccaaagtg aacttccctg | aaacaaagg ggcaggcacg gttatgctct cgtctcctgg gtgcaccacc gtcaggctga ctgggattac ggctcaagtc | 10560 10620 10680 10740 10800 10860 10920 10980 |
| tgttgcccag gttcaagcaa acacccggct tctcaaaatc aggcatgagc acacccactg tttgatattt | gtgccatagg gcaacagga catccaagga tttagctatt gctggagtgc ttctcgtgcc aattttgtat ctgacctca cactgcacct gtaaggagtt | acccatgcac cagggctgaa ttattttatt agcggcatga tcagcctcc ttttagtaga gtgatccgcc ggcctattta catggagttc ttcgtagact | tgtgaatcta gtgcctccgg tacttacttct tcttggctca aagtagctgg gatgggcttt cacctcagcc accattttaa aattccct | ggattattto gcaagggcag ctgagacaga ctgcaaccto gatttcaggg tcccaaagtg acttcctgg ttactcagga | aaaacaaagg ggcaggcacg ggtatgctct cgtctcctgg gtgcaccacc gtcaggctga ctgggattac ggctcaagtc ggttaccctcc | 10560 10620 10680 10740 10800 10860 10920 10980 11040 |
| tgttgcccag gttcaagcaa acacccggct tctcaaaatc aggcatgagc acacccactg tttgatattt gtttctgtga | gcaacagga caaccagga tttagctatt gctggagtgc ttctcgtgca aattttgtat ctgacctcag cactgcacct gtaaggagtt tctgtaattc | acccatgcac cagggctgaa ttattttatt agcggcatga tcagcctccc ttttagtaga gtgatccgcc ggcctattta catggagttc ttcgtagact tcccatggg | tgtgaatcta gtgcctccgg tacttacttt tcttggctca aagtagctgg gatgggcttt cacctcagcc accattttaa aatttcccct | ggattatite gcaagggcag ctgagacaga ctgcaaccte gatttcagge caccatgitg tcccaaagtg acttccctg ttactcagga | aaaacaaagg ggcaggcacg gttatgctct cgtctcctgg gtgcaccacc gtcaggctga ctgggattac ggtcaagtc gttaccctc ctatatcaca gaggctgcag | 10560 10620 10680 10740 10800 10860 10920 10980 11040 11100 |
| tgttgcccag gttcaagcaa acacccggct tctcaaaatc aggcatgagc acacccactg tttgatattt gtttctgtga | gcaacagga caaccagga tttagctatt gctggagtgc ttctcgtgca aattttgtat ctgacctcag cactgcacct gtaaggagtt tctgtaattc | acccatgcac cagggctgaa ttattttatt agcggcatga tcagcctccc ttttagtaga gtgatccgcc ggcctattta catggagttc ttcgtagact tcccatggg | tgtgaatcta gtgcctccgg tacttacttt tcttggctca aagtagctgg gatgggcttt cacctcagcc accattttaa aatttcccct | ggattatite gcaagggcag ctgagacaga ctgcaaccte gatttcagge caccatgitg tcccaaagtg acttccctg ttactcagga | aaaacaaagg ggcaggcacg ggtatgctct cgtctcctgg gtgcaccacc gtcaggctga ctgggattac ggctcaagtc ggttaccctcc | 10560 10620 10680 10740 10800 10860 10920 10980 11040 |
| tgttgcccag gttcaagcaa acacccggct tctcaaaatc aggcatgacgcat tttgatattt gtttctgtga gcttcaggtc | gtgccatagg gcaacaggaa catccagga tttagctatt gctggagtgc ttctcgtgcc aattttgtat ctgacctcag cactgcacct gtaaggagtt tcttgtatt | acccatgcac cagggctgaa ttatttatt agcggcatga tcagcctccc ttttagtaga gggctattta catggagttc ttcgtagac tcccatggga tgccatctg | tgtgaatcta gtgcctrccgg tacttacttt tcttggctca aagtagcttg gatgggcttt cacctcagcc accatttaa acttcccct ggggatacac cccctgcag ccagtagaaa | ggattattto gcaaggycag ctgagacaga ctgcaaccto gatttcaggo caccatgtu tcccaaagu acttccctg ttactcagga cggcagctgg | aaaacaaagg ggcaggcacg gttatgctct cgtctcctgg gtgcaccacc gtcaggctga ctgggattac ggctcaagtc gttaccctcc catattcac gaggctgcag atcaggctgcag | 10560 10620 10680 10740 10800 10860 10920 10980 11040 11100 |
| tgttgcccag gttcaagcaa acacccgcgct tctcaaaatc aggcatgagc acaccactg tttgatattt gtttctgtga gcttcaggtc gagtgtggac | gtgccatagg gcaacagga tttagctatt gctggagtgc ttctcgtgcc aattttgtat ctgacctcag cactgcacct gtaaggagtt tctgtaattc ccacctgtta ccagtggggt | acccatgcac cagggctgaa ttatttatt agcgcctcc ttttagtaga gtgatccgcc ggcctattta catggagtt ttcgtagact tcccatggga tgccatctga atctcaatga | tgtgaatcta gtgcctccgg tacttacttct tcttggctca aagtagctgg gatgggcttt cacctcagcc accatttaa aatttcccct ggggatacac cccactgcac ccagtggaaac | ggattattc gcaagggcagg ctggaacctc gatttcaggc caccatgttg tcccaaaggg aacttccctg ttactcagga cgggcagctg ctgatgtaga | aaaacaaagg ggcaggcacg gttatgctct cgtctcctgg gtgcaccacc gtcaggctga ctgggattac ggctcaagtc catattcaca gaggctgcag atcagggcgc | 10560 10620 10680 10740 10860 10920 10980 11040 11160 11160 11220 |
| tgttgcccag gttcaagcaa acacccgcgctcaaaatc aggcatgagc acacccactg tttgatattt gtttctgtga gcttctgtaggccaaggcaagg | gtgccatagg gcaacaggaa catccaagga tttagctatt gctggagtgc aattttgtat ctgacctcag caacaggagtt tctgtgcacct gtaaggagtt tctgtaattc ccacctgtta ccactgtcac ccactgtcac ccactgtcac | acccatgcac cagggctgaa ttatttatt agcggcatga tcagcctcc ttttagtaga gtgatccgcc ggcctattta catggagtt ttcgtagact tcccatggga tactctagtga actggagttgactctgc | tgtgaatcta gtgcctccgg tacttacttct tcttggctca aagtagctgg gatgggcttt cacctcagcc accatttaa aatttccct ggggatacac cccactgcag ccctcagcgc cacagtggat | ggattattc gcaagggcag ctgagacaga ctgcaacctc gatttcaggc caccatgttg tcccaaagtig tactccagga cattccttga gggcagctgg ctgatgtaga gctgaaacag | aaaacaaagg ggcaggcacg ggtatgctct cgtctcctgg gtgcaccacc gtcaggctga ctgggattac ggctcaagtc cgttacctcc catattcaca gaggctgcag atcagggcgc gtagaaatta ccaggggcag | 10560 10620 10680 10740 10860 10920 10980 11040 11160 11160 11220 11280 |
| tgttgcccag gttcaagcaa acacccgcgctcaaaatc aggcatgagc acacccactg tttgatattt gtttctgtga gcttctgtaggccaaggcaagg | gtgccatagg gcaacaggaa catccaagga tttagctatt gctggagtgc aattttgtat ctgacctcag caacaggagtt tctgtgcacct gtaaggagtt tctgtaattc ccacctgtta ccactgtcac ccactgtcac ccactgtcac | acccatgcac cagggctgaa ttatttatt agcggcatga tcagcctcc ttttagtaga gtgatccgcc ggcctattta catggagtt ttcgtagact tcccatggga tactctagtga actggagttgactctgc | tgtgaatcta gtgcctccgg tacttacttct tcttggctca aagtagctgg gatgggcttt cacctcagcc accatttaa aatttccct ggggatacac cccactgcag ccctcagcgc cacagtggat | ggattattc gcaagggcag ctgagacaga ctgcaacctc gatttcaggc caccatgttg tcccaaagtig tactccagga cattccttga gggcagctgg ctgatgtaga gctgaaacag | aaaacaaagg ggcaggcacg gttatgctct cgtctcctgg gtgcaccacc gtcaggctga ctgggattac ggctcaagtc catattcaca gaggctgcag atcagggcgc | 10560 10620 10680 10740 10860 10920 10980 11040 11160 11160 11220 |

```
actgaatcca ctgtttcatt tqttggtttg tttgttttgt tttgagaggc ggtttcactc
                                                                                                          11400
tigitgetca ggctggaggg agigcaaigg cgcgaictig gctiacigca gccicigcic cccaggitca agigaticic cigcitccgc cicccattig gcigggatia caggcacccg
                                                                                                          11460
                                                                                                          11520
ccaccatgcc cagctaattt tttgtatttt tagtagagac gggggtgggg gtggggttca
                                                                                                          11580
ccatgttggc caggctggtc tcgaacttct gacctcagat gatccacctg cctctgcctc
                                                                                                          11640
ctaaagtgct gggattacag gtgtgagcca ccatgcccag ctcagaattt actctgttta
                                                                                                          11700
                                                                                                          11760
gaaacatctg ggtctgaggt aggaagctca ccccactcaa gtgttgtggt gttttaagcc
gatacatery gyttengygt agyangted tectetted gyttyttgyf ytttaaget aatgatagaa ttrttttatt gyttyttagaa cactettgat gytttacaet gygatgatgata agacateate agctttteaa agacacaeta actgeaceca taatactggg gygtettetg gytateageg atetteattg aatgeeggga gycgttteet cyceatycae atgytyttaa ttactecage ataatettet getteeattt ettetettee etetttaaa attytyttt
                                                                                                          11820
                                                                                                           11880
                                                                                                           11940
                                                                                                           12000
ctatottogc ttctctgcag agaaccagtg taagctacaa cttaactttt gttggaacaa
                                                                                                           12060
attttccaaa ccgccccttt gccctagtgg cagagacaat tcacaaacac agccctttaa
                                                                                                           12120
aaaggettag ggateactaa ggggatttet agaagagega eeegtaatee taagtattta
                                                                                                          12180
caagacgagg ctaacctcca gcgagcgtga cagcccaggg agggtgcgag gcctgttcaa
                                                                                                          12240
                                                                                                          12300
atgctagete cataaataaa gcaattteet eeggeagttt etgaaagtag gaaaggttae
atttaaggtt gcgtttgtta gcatttcagt gtttgccgac ctcagctaca gcatccctgc
                                                                                                           12360
aaggcctcgg gagacccaga agtttctcgc cccttagatc caaacttgag caacccggag
tctggattcc tgggaagtcc tcagctgtcc tgcggttgtg ccggggcccc aggtctggag
                                                                                                           12420
                                                                                                           12480
gggaccagtg gccgtgtggc ttctactgct gggctggaag tcggggctct tagctctgga
gtccgaggct tggagccagg tgcctggacc ccgaggctgc cctccaccct gtgcgggcgg
                                                                                                           12540
                                                                                                           12600
gatgtgacca gatgttggcc tcatctgcca gacagagtgc cggggcccag ggtcaaggcc
                                                                                                           12660
gttgtggctg gtgtgaggcg cccggtgcgc ggccagcagg agcgcctggc tccatttccc
                                                                                                           12720
accettlete gacgggaccg ceceggtggg tgattaacag atttggggtg gtttgeteat
                                                                                                          12780
                                                                                                           12840
ggtggggacc cctcgccgcc tgagaacctg caaagagaaa tgacgggcct gtgtcaagga
                                                                                                           12900
gcccaagtcg cggggaagtg ttgcagggag gcactccggg aggtcccgcg tgcccgtcca
gcccadgrcy cygggadyt regraygyay gcarregyy aggreecyt tycetyrea
gggagcaatg cgtcctcggg tregtccca gccgcgtcta cgcgcctcc gaggcagcc
cacgtccggc attegtggtg cccggagcc gacgcccgc gtccggacct ggaggcagcc
ctgggtctcc ggatcaggcc agcggccaaa gggtcgccgc acgcacctgt tcccagggcc
tccacatcat ggcccctccc tcgggttacc ccacagccta ggccgattcg acctctctcc
                                                                                                           12960
                                                                                                           13020
                                                                                                           13080
                                                                                                           13140
gctggggccc tcgctggcgt ccctgcaccc tgggagcgcg agcggcgcgc gggcggggaa
                                                                                                           13200
gcgcggccca gacccccggg tccgcccgga gcagctgcgc tgtcggggcc aggccgggct
                                                                                                           13260
cccagtggat tcgcgggcac agacgcccag gaccgcgctt cccacgtggc ggagggactg
                                                                                                          13320
gggacccggg cacccgtcct gccccttcac cttccagctc cgcctcctcc gcgcggaccc
                                                                                                           13380
 egececgice egacecetee egggiceeeg geceageeee eteegggeee teeeageeee
                                                                                                           13440
tccccttcct ttccgcggcc ccgccctctc ctgcggcgc gagtttcagg cacgccttgcc
tcctgctgcg cacgtgggaa gccctggccc cggccacccc cgcgatgccg cgcgctccc
gctgccgagc cgtgcgctc ctgctgcgca gccactaccg cgaggtgctg ccgctggcca
cgttcgtgcg gcgcctgggg ccccagggct ggcggctggt gcagcgcggg gacccggcgg
                                                                                                           13500
                                                                                                           13560
                                                                                                           13620
                                                                                                           13680
ctttcggcg gctggtggcc cagtgctgg tgtggtgcc ctggggacgca cggccgccc ccgccgcccc ctccttccgc cagtggcgc tccccggggt cggcgtccg ctggggttga
                                                                                                           13740
                                                                                                           13800
gggcggccgg ggggaaccag cgacatgcgg agagcagcgc aggcgactca gggcgcttcc
                                                                                                           13860
 cccgcaggtg tcctgcctga aggagctggt ggcccgagtg ctgcagaggc tgtgcgagcg
                                                                                                           13920
cggcgcgaag aacgtgctga aggattggt cgccgctgt gacggggcccccccccgaggcgaag acgtgctgaggctgcag ctactcacca gcgtgcgcag ctacttgccc aacacggtga ccgacgcact gcgggggagc ggggggcgtggg ggctgctgct gcgccgcgtg ggcgacgacg tgctggttca cctgctggca cgctgcgcc tctttgtgct ggtggctccc agctgcgcct accaggtgtg cgggccgccg ctgtaccagc tcggcgctgc cactcaggcc cggcccccg cacacgctag
                                                                                                           13980
                                                                                                           14040
                                                                                                           14100
                                                                                                           14160
                                                                                                           14220
 tggaccccga aggcgtctgg gatgcgaacg ggcctggaac catagcgtca gggaggccgg
                                                                                                           14280
ggtcccctg ggcctgccag ccccgggtgc gaggaggcgc gggggcagtg ccagccgaag
                                                                                                           14340
 tctgccgttg cccaagaggc ccaggcgtgg cgctgcccct gagccggagc ggacgcccgt
                                                                                                           14400
                                                                                                           14460
 tgggcagggg tcctgggccc acccgggcag gacgcgtgga ccgagtgacc gtggtttctg
                                                                                                           14520
 tgtggtgtca cctgccagac ccgccgaaga agccacctct ttggagggtg cgctctctgg
cacgogocae teccaccat cogtogocog cagcaccae gogogococc catecacate gogogocacca egtectogog acacgoctty tecceogoty taegeogaga ceaagcaett ectetactee teaggogaca aggageaget gogococtee tectactea getetetgag geceagectg actggogote ggaggetegt ggagaccate tetetgggtt ceaggecetg
                                                                                                           14580
                                                                                                           14640
                                                                                                           14700
                                                                                                           14760
gatgccaggg actccccgca ggttgccccg cctgccccag cgctactggc aaatgcggcc
                                                                                                           14820
 cctgtttctg gagctgcttg ggaaccacgc gcagtgcccc tacggggtgc tcctcaagac
                                                                                                           14880
 gcactgcccg ctgcgagctg cggtcacccc agcagccggt gtctgtgccc gggagaagcc
                                                                                                           14940
 ccagggctct gtggcggccc ccgaggagga ggacacagac ccccgtcgcc tggtgcagct
                                                                                                           35000
                                                                                                           15060
gctccgccag cacagcagcc cctggcaggt gtacggcttc gtgcgggcct gcctgcgccg
gctcgccac caagcagcc ctrygcagy glatygcttc grycygytt gctgryctg
gctggtgccc caagcagcctct ggggctccag gcacaacgaa cgccgcttcc tcaagaacac
caagaagttc atctccctgg ggaagcatgc caagctctcg ctgcaggagc tgacgtggaa
gatgagcgtg cgggactgcg cttggctgcg caggagccca ggtgaggagg tggtggccgt
cgagggccca ggccccagag ctgaatgcag taggggctca gaaaaggggg caggcagagc
                                                                                                           15120
                                                                                                           15180
                                                                                                           15240
                                                                                                           15300
cctggtctc ctgtctccat cgtcacgtgg gcacacgtgg cttttcgctc aggacgtcga
gtggacacgg tgatcgagtc gactcccttt agtgagggtt aattgagctc gcggccgc
                                                                                                           15360
                                                                                                           15418
```

SEQ. ID NO:2 (human blood group A transferase)

```
1 atggccgagg tgttgcggac gctggccgga aaaccaaaat gccacgcact tcgacctatg 61 atccttttcc taataatgct tgtcttggtc ttgtttggtt acggggtcct aagccccaga 121 agcctaatgc caggaagcct ggaacggggg ttctgcatgg ctgttaggga acctgaccas 181 ctgcagcgcg tctcgttgcc aaggatggtc tacccccagc caaaggtgct gacaccgtgg 241 aaggatgtcc tcgtggtgac cccttggctg gctcccattg tctgggaggg cacattcaac 301 atcgacatcc tcaacgagca gttcaggctc cagaacacca ccattgggtt aactgtgtt 361 gccatcaaga aatacgtggc ttcctgaag ctgttcctgg agacggcgga gaagcacttc 421 atggtggggc accggtcca ctactatgt ttcaccgacc agctggccgc ggtgccccgc 481 gtgacgctgg gaaccggtcg gcactggcag acatggacgt ccatggaggt ccatggaggt ccatggaggt ccatggaggat acagtgact tctgcgagcg gcgcttcctc
```

```
601 agcgaggtgg attacctggt gtgcgtggac gtggacatgg agttccgcga ccacgtgggc 661 gtggagatcc tgactccgct gttcggcacc ctgcaccccg gcttctacgg aagcagcggg 721 gaggccttca cctacgagcg ccggccccag tcccaggcct acatccccaa ggacgagggg 781 gatttctact acctggggg gttcttcggg gggtcggtgc aagagggtga gcggctacc 841 agggcctgcc accaggccat gatggtcgac caggccaacg gcatcgaggc cgtgtggcac 901 gacgagagcc acctgaacaa gtacctgctg cgccacaaac ccacaaggt gcttccccc 961 gagtacttgt gggaccagca gctgcgggc tggccgccg tcctgaggaa gctgaggttc 1021 actgcggtgc ccaagaacca ccaggcggtc cggaacccgt ga
```

SEQ. ID NO:3 (human blood group A transferase)

MAEVLRTLAGKPKCHALRPMILFLIMLVLVLFGYGVLSPRSLMP GSLERGFCMAVREPDHLQRVSLPRMVYPQPKVLTPWKDVLVVTPWLAPIVWEGTFNID ILNEQFRLQNTTIGLTVFAIKKYVAFLKLFLETAEKHFMVGHRVHYYVFTDQLAAVPR VTLGTGRQLSVLEVRAYKRWQDVSMRRMEMISDFCERRFLSEVDYLVCVDVDMEFRDH VGVEILTPLFGTLHPGFYSGSREAFTYERRPQSQAYIPKDEGDFYYLGGFFGGSVQEV QRLTRACHQAMMVDQANGIEAVWHDESHLNKYLLRHKPTKVLSPEYLWDQQLLGWPAV LRKLRFTAVPKNHOAVRNP

SEQ. ID NO:4 (human blood group B transferase)

SEQ. ID NO:5 (hmman blood group B transferase)

MAEVLRTLAGKPKCHALRPMILFLIMLVLVLFGYGVLSPRSLMP GSLERGFCMAVREPDHLQRVSLPRMVYPQPKVLTPCRKDVLVVTPWLAPIVWEGTFNI DILNEQFRLQNTTIGLTVFAIKKYVAFLKLFLETAEKHFMVGHRVHYYVFTDQPAAVP RVTLGTGRQLSVLEVGAYKRWQDVSMRMEMISDFCERRFLSEVDYLVCVDVDMEFRD HVGVEILTPLFGTLHPSFYGSSREAFTYERRPQSQAYIPKDEGDFYYMGAFFGGSVQE VQRLTRACHQAMMVDQANGIEAVWHDESHLNKYLLRHKPTKVLSPEYLWDQQLLGWPA VLRKLRFTAVPKNHQAVRNP

SEQ. ID NO:6 (marmoset al,3GT)

MNVKGKVILSMLVVSTVIVVFWEYINSPEGSFLWIYHSKNPEV-DDSSAQKDWWFPGWFNNGIHNYQQEE EDTDK-EKGREEEQKKEDDTTELRLWDWFNPKKRPEVMTVTQWKAPVVWEGTYNKAILENYYAKQKITVG LTVFAIGRYIEHYLEEFVTSANRYFMVGHKVIFYVMVDDVSKAPFIELGPLRSFKVFEVKPEKRWQDISM MRMKTIGEHILAHIQHEVDFLFCMDVDQVFQDHFGVETLGQSVAQLQAWWYKADPDDFTYERRKESAAYI PFCQGDFYYHAAIFGGTPIQVLNITQECFKGILLDKKNDIEAEWHDESHLNKYFLLNKPSKILSPEYCWD YHIGLPSDIKTVKLSWQTKEYNLVRKNV

SEQ. ID NO:7 (human α1,3GT pseudogene)

RYNDHYLEEFITSANRYFMVGHKVIFYIMVDDVSKLPFIELGPLHSFKMFEVKPEKRWQDISM MRMKITGEHILAHIQHEVDFLFCMDVDQVFQDHFGVETLGQSVAQLQ*WRYKADPYDFT*ERWKESAGYI PFG*GDFYYHAAISGGTPIQVLNITQECFKGILLDKKNDIEAKWHDESHLNKYFLLNKPSKILSLKYCWD YHIGLPSDIKTVK*SWQTKEYNLVRNNV

SEQ. ID NO:8 (sheep α 1,3GT)

MNVKGKVILSMLVVSTVIVVFWEYIHSPEGSLFWINPSRNPEVSGGSSIQKGWWFPRWFNNG---Y-QEE
DEDVDEEKEQRKEDK-----SKLKLSDWFNPFKRPEVVTMTDWKAPVVWEGTYNRAVLDDYYAKQKITVG
LTVFAVGRYIEHYLEEFLTSANKHFMVCHRVIFYVMVDDVSRMPLIELGPLRSFKVFEVKPERRWQDVSM

VRMKTIGEHIVAHIQREVDFLFCMDVDQVFQDEFGVETLGESVAQLQAWWYKADPDEFTYERRKESAAYI PFGEGDFYYHAAIFGCTPTQVLNITQECFKGILKDKKNDIEAQWHDESHLNKYFLLNKPTKILSPEYCWD YHIGLPADIKLVKMSWQTKEYNVVRNNV

SEQ. ID NO:9 (bovine $\alpha1,30T$)

MNVKGKVILSMLVVSTVIVVFWEYIHSPEGSLFWINPSRNPEV-GGSSIQKGWWLPRWFNNG---Y-HEE DGDINEEK----EQRNEDE-SKLKLSDWFNPFKRPEVVTMTKWKAPVVWEGTYNRAVLDNYYAKQKITVG LTVFAVGRYIEHYLEEFLTSANKHFMVGHPVIFYIMVDDVSRMPLIELGPLRSFKVFKIKPEKRWQDISM MRMKTIGEHIVAHIQHEVDFLFCMDVDQVFQDKFGVETLGESVAQLQAWWYKADPNDFTYERRKESAAYI PFGEGDFYYHAAIFGGTPTQVLNITQECFKGILKDKKNDIEAQWHDESHLNKYFLLNKPTKILSPEYCWD YHIGLPADIKLYKMSWQTKEYNVVRNNV

SEQ. ID NO:10 (pig a1,3GT)

MNVKGRVVLSMLLVSTVMVVFWEYINSPEGSLFWIYQSKNPEV--GSSAQRGWWFPSWFNNGTHSY-HEE EDAIGNEK----EQRKEDNRGELPLVDWFNPEKRPEVVTITRWKAPVVWEGTYNRAVLDNYYAKQKITVG LTVFAVGRYIEHYLEEFLISANTYFMVGHKVIFYJMVDDISRMFNPLIELGPLRSFKVFEIKYSKKRÜDISM MRMKTIGEHILAHIQHEVDFLFCMDVDQVFQNNFGVETLGQSVAQLQAWYKAHPDEFTYERRKESAAYI PFGQGDFYYHAAIFGGTPTQVLNITQECFKGILQDKENDIEAEWHDESHLNKYFLLNKPTKILSPEYCWD YHIGMSVDIRIVKIAWQKKEYNLVRNNI

SEQ. ID NO:11 (mouse α1,3GT)

MNVKGKVILLMLIVSTVVVVFWEYVNRI------PEV-GENRWQKDWWFPSWFKNGTHSY-QED NVEGRREK----GRNGDRIEEPQLWDWFNPKNRPDVLTVTPWKAPIVWEGTYDTALLEKYYATQKLTVG LTVFAVGKYIEHYLEDFLESADMYFMVGHRVIFYYMIDDTSRMPVVHLNPLHSLQVFEIRSEKRWQDISM MRMKTIGEHILAHIQHEVDFLFCMDDQVFQDNFGVETLGQLVAQLQAWWYKASPEKFTYERRELSAAYI PFGEGDFYYHAAIFGGTPTHILNLTRECFKGILQDKKHDIEAQWHDESHLNKYFLFNKPTKILSPEYCWD YQIGLPSDIKSVKVAWQTKEYNLVRNNV

SEQ. ID NO:12 (consensus al,3GT)

MNVKGKVILSMLVVSTVIVVFWEYINSPEGSFLWIYHSKNPEV-DDSSAQKDWFPGWFNNGTHNYQQEE EDTDK-EKGREEEQKKEDDTTELRLWDWFNPKKRPEVMTVTQWKAPVVWEGTYNKAILENYYAKQKITVG LTVFAIGRYIEHYLEEFLTSANRYFMYGHKVIFYYMVDDVSKAPFIELGPLRSFKVFEVKPEKRWQDISM MRMKTIGEHILAHIQHEVDFLFCMDVDQVFQDHFGVETLGQSVAQLQAWWYKADPDDFTYERKESAAYI PFGQGDFYYHAAIFGGTPIQVLNITQECFKGILLDKKNDIEAEWHDESHLNKYFLLNKPSKILSPEYCWD YHIGLPSDIKTVKLSWQTKEYNLVRKNV

SEQ. ID NO:13 (humanized α1,3GT)

MNVKGKVILSMLVVSTVIVVFWEYINSPEGSFLWIYHSKNPEV-DDSSAQKDWWFPGWFNNGIHNYQQEE EDTDK-EKGREEEQKKEDDTTELRLWDWFNPKKRPEVMTVTQWKAPVVWEGTYNKAILENYYAKQKITVG LTVFAIGRYIDHYLEEFLTSANRYFMVGHKVIFYIMVDDVSKAPFIELGPLRSFKVFEVKPEKRWQDISM MRMKITGEHILAHIQHEVDFLFCMDVDQVFQDHFGVETLGQSVAQLQAWWYKADPDDFTYERRKESAGYI PFGQGDFYYHAAIFGGTPIQVLNITQECFKGILLDKKNDIEAEWHDESHLNKYFLLNKPSKILSPEYCWD YHIGLPSDIKTVKLSWQTKEYNLVRKNV

SEQ. ID NO:14 (marmoset α1,3GT)

| 1 | atgaatgtcaaaggaaaagtaattctgtcgatgctggttgtctcaactgtgattgttgtg | |
|------|--|------|
| 61 | ttttgggaatatatcaacagcccagaaggctctttcttgtggatatatcactcaaagaac | |
| 121 | ccagaagttgatgacagcagtgctcagaaggactggtggtttcctggctgg | |
| 181 | gggatccacaattatcaacaagaggaagaagacacagacaaagaaaaaggaagagagag | |
| 241 | gaacaaaaaaaggaagatgacacaacagagcttcggctatgggactggtttaatccaaag | |
| 301 | aaacgcccagaggttatgacagtgacccaatggaaggcgccggttgtgtgtg | |
| 361 | tacaacaaagccatcctagaaaattattatgccaaacagaaaattaccgtggggttgacg | |
| 421 | gtttttgctattgga | |
| 436 | agatatattgagcattacttggaggagttcgtaacatctgctaataggtacttcatggtc | 495 |
| 496 | ggccacaaagtcatattttatgtcatggtggatgatgtctccaaggcgccgtttatagag | 555 |
| 556 | ctgggtcctctgcgttccttcaaagtgtttgaggtcaagccagagaagaggtggcaagac | 615 |
| 616 | atcagcatgatgcgtatgaagaccatcggggagcacatcttggcccacatccaacacgag | 675 |
| 676 | gttgacttcctcttctgcatggatgtggaccaggtcttccaagaccattttggggtagag | 735 |
| 736 | accctgggccagtcggtggctcagctacaggcctggtggtacaaggcagatcctgatgac | 795 |
| 796 | tttacctatgagaggcggaaagagtcggcagcatatattccatttggccagggggatttt | 855 |
| 856 | tattaccatgcagccatttttggaggaacaccgattcaggttctcaacatcacccaggag | 915 |
| 916 | tgctttaagggaatcctcctggacaagaaaaatgacatagaagccgagtggcatgatgaa | 975 |
| 976 | agccacctaaacaagtatttccttctcaacaaaccctctaaaatcttatctccagaatac | 1035 |
| 1036 | tgctgggattatcatataggcctgccttcagatattaaaactgtcaagctatcatggcaa | 1095 |
| 1096 | acaaaagagtataatttggttagaaagaatgtctga | 1131 |
| | | |

PCT/US01/44306 WO 02/42468

SEQ. ID NO:15 (human α1,3GT pseudogene)

cagcttgtggtttctttcaggaatcccagaggataaatgttttgcttttcttctttgtttcagatataatgatcattacttggaggaggttcataacatctgctaataggtacttcatggtt ggccacaaagtcatattttacatcatggtggatgatgtctccaagctgccgtttatagag 62 121 181 122 ctgggtcctctgcattccttcaaaatgtttgaggtcaagccagagaagaggtggcaagac atcagcatgatgcgtatgaagatcactggggagcacatcttggcccacatccaacacgag gtcgacttcctcttctgcatggatgtggaccaggtcttccaagaccattttgggggtggag 241 182 242 301 302 361 accctaggccagtcagtggctcagctacagg-ctggcggtacaaggcagatccctatgac 420 362 tttacctaggagaggtggaaagagtcagcaggatacattccatttggcca-ggggatttt 479 421 480 tattaccatgcagccatttctggaggaacacccattcaggttctcaacatcacccaggag 539 540 tgctttaagggaatcctcctggacaagaaaaatgacatagaagccaagtggcatgatgaa 599 600 agccacctaaacaagtatttccttctcaataaaccctctaaaatcttatccctaaaatac 659 660 tgctgggattatcatataggcctgccttcagatattaaaactgtcaagtgatcgtggcag 719 755 720 acaaaagagtataatttggttagaaataatgtctga

SEQ. ID NO:16 (Humanized α1,3GT)

ttttgggaatatatcaacagcccagaaggctctttcttgtggatatatcactcaaagaac 181 gggatccacaattatcaacaagaggaagaagacacagacaaagaaaaaggaagagagag 241 gaacaaaaaaaggaagatgacacaacagagcttcggctatgggactggtttaatccaaag 301 aaacgcccagaggttatgacagtgacccaatggaaggcgccggttgtgtggggaaggcact 361 tacaacaaagccatcctagaaaattattatgccaaacagaaaattaccgtggggttgacg 421 gtttttgctattgga agatatattgatcattacttggaggagttcttaacatctgctaataggtacttcatggtt 436 ggccacaaagtcatattttacatcatggtggatgatgtctccaaggcgccgtttatagag 496 ctgggtcctctgcgttccttcaaagtgtttgaggtcaagccagagaagaggtggcaagac 556 atcagcatgatgcgtatgaagatcactggggagcacatcttggcccacatccaacacgag 675 676 gtcgacttcctcttctgcatggatgtggaccaggtcttccaagaccattttggggtggag 735 736 accctaggccagtcagtggctcagctacaggcctggtggtacaaggcagatcccgatgac 795 796 $\verb|tttacctatgagaggcggaaagagtcagcaggatacattccatttggccagggggatttt|$ 855 915 856 tattaccatgcagccatttttggaggaacacccattcaggttctcaacatcacccaggag tgctttaagggaatcctcctggacaagaaaaatgacatagaagccgagtggcatgatgaaagccacctaaacaagtatttccttctcaataaaccctctaaaatcttatcccagaatac 975 916 976 1035 tgctgggattatcatataggcctgccttcagatattaaaactgtcaagctatcgtggcag 1036 1096 acaaaagagtataatttggttagaaataatgtctga

SEQ. ID NO:17 (Sheep a1,3GT)

agccgaggacgccgcggggagccgaggctccggccagccccagcgcgcccagcttctgcagatcagg

agtcagaacgctgcac

cttcgcttcctccagccctgcctccttctgcaaaacggagctcaatagaacttggtact tttgccttttactctgggaggagagaagcagacgatgaggagaaaata

[beginning of coding sequence]

atgaatgtcaaa

ggaaaagtgattctgtcaatgctggttgtctcaactgtcattgttgtgttttgggaatat atccacageccagaaggctctttgttctggataaacccatcaagaaacccagaagtcagt aagctatcggactggttcaacccatttaaacgccctgaggttgtgactatgacagattgg aaggcacccgtggtgtggggaaggcacttacaacagagccgtcttagacgattactacgcc aagcagaaaattaccgtcggcctgacggttttcgccgtcggaagatacattgagcattac ttggaggagttcttaacgtctgctaataagcacttcatggttggccaccgagtcatcttt tacgtcatggtggacgacgtctccaggatgcctttgatagagctgggccctctgcgctccttcaaagtgtttgaggtcaagcctgagaggaggtggcaggacgtcagcatggtgcgcatg aagaccatcggggagcacatcgtggcccacatccagcgtgaggttgacttcctettctgc atggacgtggaccaggtcttccaagacgagttcggggtggagaccctgggtgagtcggtg gcccagctacaggcctggtggtacaaggcagatcccgatgagtttacctacgagaggcgc aaggagtctgcagcatacattcccttcggcgaaggggatttttattaccacgcagccatt tttgggggaacacccactcaggtccttaacatcacccaggaatgcttcaaaggaatcctc aaggacaagaaaaatgacatagaagcccaatggcatgatgagagccatctaaacaagtat ttccttctcaacaaacccactaaaatcttatccccggaatactgctgggattatcatata ggcctacctgcggatattaagcttgtcaagatgtcttggcagacaaaagagtataatgtg gttagaaataacgtctga [end of coding sequence]

CLAIMS

What is claimed as the invention is:

- A polynucleotide comprising an encoding sequence for a glycosyltransferase under control of a
 heterologous tumor specific or tissue specific transcriptional control element, wherein expression
 of the polynucleotide in a human cell causes the cell to express a cell-surface carbohydrate
 determinant to which some or all humans have naturally occurring antibody.
- 2. The polynucleotide of claim 1, wherein the glycosyltransferase is a blood group A transferase.
- 3. The polynucleotide of claim 1, wherein the glycosyltransferase is a blood group B transferase.
- 4. The polynucleotide of claim 1, wherein the glycosyltransferase is an $\alpha(1,3)$ galactosyltransferase.
- 5. The polynucleotide of claim 1, wherein the encoding sequence encodes either:
 - a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ.
 ID NOs:3, 5, 6, 12, and 13; or
 - a fragment thereof having galactosyltransferase activity.
- 6. A polynucleotide comprising an encoding sequence for a human ABO histo blood group transferase under control of a tissue or tumor specific transcriptional control element.
- 7. The polynucleotide of claims 1-6, wherein the transcriptional control element is a tissue specific promoter, which is a promoter for albumin, α-fetoprotein, prostate-specific antigen (PSA), mitochondrial creatine kinase (MCK), myelin basic protein (MB), glial fibrillary acidic protein (GFAP), or neuron-specific enolase (NSE).
- 8. The polynucleotide of claims 1-6, wherein the transcriptional control element is a tumor specific promoter, which is a promoter for telomerase reverse transcriptase (TERT), carcinoembryonic antigen (CEA), hypoxia-responsive element (HRE), *Grp78*, L-plastin, or hexokinase II.
- 9. The polynucleotide of claim 8, wherein the promoter comprises at least 25 consecutive nucleotides in SEQ. ID NO:1.
- 10. A viral vector comprising the polynucleotide of any preceding claim.
- 11. The vector of claim 10, which is an adenovirus vector.
- A humanized or consensus α(1,3)galactosyltransferase, comprising the amino acid sequence shown in SEQ. ID NOs:12 or 13, or a fragment thereof having galactosyltransferase activity.

13. A method of killing a cancer cell, comprising combining the cancer cell with the polynucleotide of any of claims 1-11.

- 14. A human cell containing the polynucleotide or vector of any of claims 1-11.
- 15. A medicament comprising the polynucleotide of any of claims 1-11.
- 16. Use of a polynucleotide according to any of claims 1-11 in the preparation of a medicament for the treatment of cancer.

Figure 1

pGRN376

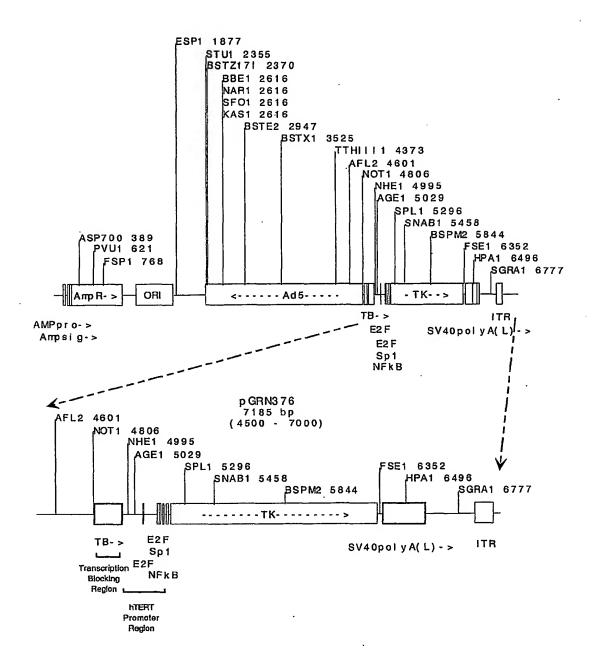
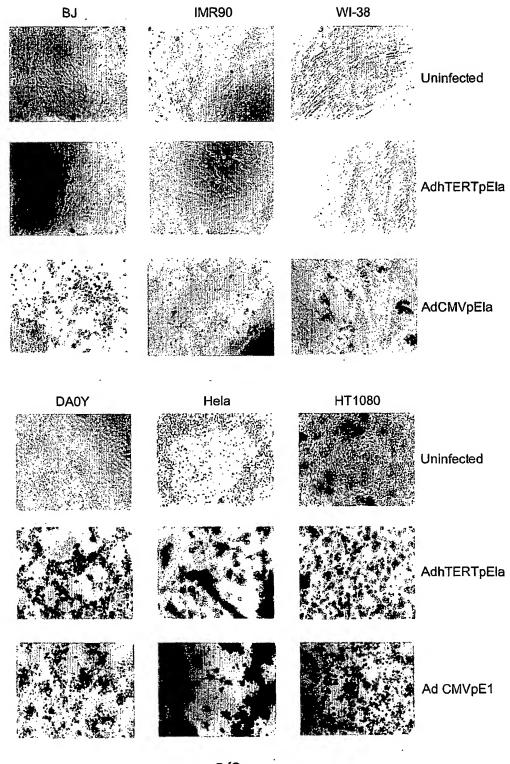


Figure 2



2/6

Figure 3(A)

| Marmoset α1,3GT human pseudogene sheep Bovine Pig Mouse | MNVKGKVILSMLVVSTVIVVFWEYINSPEGSFLWIYHSKNPEV-DDSSAQKDWWFPGWFNNGIHNYQQEE |
|---|--|
| Consensus α 1,3GT Humanized α 1,3GT | |
| hu B transferase hu A transferase | .AEVLRTLAGKPKCHALRPMILFL.MLVLVL.GYGVLSPRSLMPGSL .AEVLRTLAGKPKCHALRPMILFL.MLVLVL.GYGVLSPRSLMPGSL |
| Marmoset α1,3GT human pseudogene sheep Bovine Pig Mouse Consensus α1,3GT Humanized α1,3GT hu B transferase hu A transferase | EDTDK-EKGREEEQKKEDDTTELRLWDWFNPKKRPEVMTVTQWKAPVVWEGTYNKAILENYYAKQKITVG DEDVDE. EQRK.DSK.K.SFV.M.DR.V.DD DGDINERN.E-SK.K.SFV.M.KR.V.D AIGNR.NRG.P.V.E.V.I.R.R.V.D NVEGRRGRNG.RIE.PQN.D.L.P.I.DT.L.KT.L V.M.P.I.R.V.D RGFCMAVREPDHLQRVSLPRMVYPQPKVLTPC.KD.LV.P.L.I.F.ID.NEQFRL.NT.I. RGFCMAVREPDHLQRVSLPRMVYPQPKVLTPC.KD.LV.P.L.I.F.ID.NEQFRL.NT.I. |
| Marmoset α1,3GT human pseudogene sheep Bovine Pig Mouse | LTVFAIGRYIEHYLEEFVTSANRYFMVGHKVIFYVMVDDVSKAPFIELGPLRSFKVFEVKPEKRWQDISM .ND I L H M .V L KH R RM L KI .V L KH P I RM L KI .V L I I I RM I I S .V L R I RM I I S V |
| Consensus α1,3GT Humanized α1,3GT | DLI |
| hu B transferase hu A transferase | KK.VA-F.KL.LET.EKHR.HYFT.QPAAV.RVTTG.QLS.LGAYV KK.VA-F.KL.LET.EKHR.HYFT.QLAAV.RVTTG.QLS.LRAYV |

Figure 3(B)

| Marmoset $\alpha 1,3GT$ human pseudogene sheep Bovine Pig Mouse | MRMKTIGEHILAHIQHEVDFLFCMDVDQVFQDHFGVETLGQSVAQLQAWWYKADPDDFTYERRKESAAYI IT |
|--|--|
| Consensus α 1,3GT Humanized α 1,3GT | ITE |
| hu B transferase hu A transferase | REM.SDFCERRFLSY.V.VME.RVI.TPLFGT.HPSF.GSSREAPQ.Q REM.SDFCERRFLSY.V.VME.RVI.TPLFGT.HPGF.GSSREAPQ.Q |
| Marmoset α1,3GT human pseudogene sheep Bovine Pig Mouse Consensus α1,3GT Humanized α1,3GT hu B transferase hu A transferase | PFGQGDFYYHAAIFGGTPIQVLNITQECFKGILLDKKNDIEAEWHDESHLNKYFLLNKPSKILSPEYCWD .* < |
| Marmoset α1,3GT human pseudogene sheep Bovine Pig Mouse | YHIGLPSDIKTVKLSWQTKEYNLVRKNV |
| Consensus α 1,3GT Humanized α 1,3GT | |
| hu B transferase hu A transferase | QQLLGWPAVLRKLRFTAVPKNHQAVR.P QQLLGWPAVLRKLRFTAVPKNHQAVR.P |

Figure 4(A)

| | 1 61 121 181 241 301 361 421 | atgaatgtca aaggaaaagt aattctgtcg atgctggttg tctcaactgt gattgt ttttgggaat atatcaacag cccagaaggc tctttcttgt ggataatca ctcaaag ccagaagttg atgacagcag tgctcagaag gactggtggt ttcctggctg gtttaa gggatccaca attatcaaca agaggaagaa gacacagaca aagaaaaagg aagaga gaacaaaaaa aggaagatga cacaaacagag cttcggctat gggactggtt taatcca aaacgcccag aggttatgac agtgacccaa tggaaggcgc cggttgtgtg ggaaggg tacaacaaaag ccatcctaga aaattattat gccaaacaga aaattaccgt ggggttg gtttttgcta ttgga | gaac caat ggag aaag cact |
|-------------------|---|--|--------------------------------------|
| larmoset α1,3GT: | 43 6 | agatatattgagcattacttggaggagttcgtaacatctgctaataggtacttcatggtcatt | 495 |
| luman pseudogene: | 62 | | 121 |
| lumanized α1,3GT: | 436 | t agatatattgatcattacttggaggagttcttaacatctgctaataggtacttcatggtt | 121 |
| : | 436 | | 495 |
| narmoset α1,3GT: | 496 | ggccacaaagtcatattttatgtcatggtggatgatgtctccaaggcgccgtttatagagcact | 555 |
| Numan pseudogene: | 122 | | 181 |
| lumanized α1,3GT: | 496 | ggccacaaagtcatattttacatcatggtggatgatgtctccaaggcgccgtttatagag | 181 |
| : | 496 | | 555 |
| Aarmoset α1,3GT: | 556 | ctgggtcctctgcgttccttcaaagtgtttgaggtcaagccagagaagaggtggcaagac | 615 |
| Human pseudogene: | 182 | | 241 |
| dumanized α1,3GT: | 556 | ctgggtcctctgcgttccttcaaagtgtttgaggtcaagccagagaagaggtggcaagac | 241 |
| : | 556 | | 615 |
| Marmoset α1,3GT: | 616 | atcagcatgatgcgtatgaagaccatcggggagcacatcttggcccacatccaacacgagtct | 675 |
| Human pseudogene: | 242 | | 301 |
| Humanized α1,3GT: | 616 | atcagcatgatgcgtatgaagatcactggggagcacatcttggcccacatccaacacgag | 301 |
| : | 616 | | 675 |
| Marmoset α1,3GT: | 676 | gttgacttcctcttctgcatggatgtggaccaggtcttccaagaccattttggggtagag | 735 |
| Human pseudogene: | 302 | | 361 |
| Humanized α1,3GT: | 676 676 | cggtcgacttcctcttctgcatggatgtggaccaggtcttccaagaccattttggggtggag | 361 735 |
| Marmoset α1,3GT: | 736 | accctgggccagtcggtggctcagctacaggcctggtggtacaaggcagatcctgatgacaact | 795 |
| Human pseudogene; | 362 | | 420 |
| Humanized α1,3GT: | 736 736 | aa | 420 795 |

Figure 4(B)

| Marmoset α1,3GT: | 796 | tttacctatgagaggcggaaagagtcggcagcatatattccatttggccagggggattttgtagc | 855 |
|---------------------------------------|--------------|---|-------------|
| Human pseudogene: | 421 | | 479 |
| dumanized α1,3GT: | 796 | tttacctatgagaggcggaaagagtcagcaggatacattccatttggccagggggatttt | 479 |
| : | 796 | | 855 |
| Marmoset α1,3GT: | 856 | tattaccatgcagccatttttggaggaacaccgattcaggttctcaacatcacccaggag | 915 |
| Human pseudogene: | 480 | | 539 |
| Humanized α 1,3GT: | 856 856 | tattaccatgcagccatttttggaggaacacccattcaggttctcaacatcacccaggag | 539 915 |
| Marmoset α1,3GT: | 916 | tgctttaagggaatcctcctggacaagaaaaatgacatagaagccgagtggcatgatgaa | 975 |
| Human pseudogene: | 540 | | 599 |
| Humanized α1,3GT: | 916 916 | tgctttaagggaatcctcctggacaagaaaaatgacatagaagccgagtggcatgatgaa | 599 975 |
| Marmoset α1,3GT: | 976 | agccacctaaacaagtatttccttctcaacaaaccctctaaaatcttatctccagaatac | 1035 |
| Human pseudogene: | 600 | | 659 |
| Humanized α1,3GT: | 976 976 | agccacctaaacaagtatttccttctcaataaaccctctaaaatcttatccccagaatac | 659 1035 |
| Marmoset α1,3GT: | 1036 | tgctgggattatcatataggcctgccttcagatattaaaactgtcaagctatcatggcaa | 1095 |
| Human pseudogene: | 660 | | 719 |
| Humanized α 1,3GT: | 1036 1036 | tgctgggattatcatataggcctgccttcagatattaaaactgtcaagctatcgtggcag | 719 1095 |
| Marmoset α1,3GT: Human pseudogene: | 1096 720 | acaaaagagtataatttggttagaaagaatgtctga 1131 | |
| Humanized α1,3GT: | 1096 1096 | | |

SEQUENCE LISTING

```
<110> Geron Corporation
       Schiff, J. Michael
<120> GLYCOSYLTRANSFERASE VECTORS FOR TREATING CANCER
<130> 083,002
<140>
       [to be assigned]
<141>
       2001-11-26
<150> 60/253.395
       2000-11-27
<151>
<160> 17
<170> PatentIn version 3.1
<210>
<211> 15418
<212>
       DNA
      Homo sapiens
<213>
<400> 1
gcggccgcga gctctaatac gactcactat agggcgtcga ctcgatcaat ggaagatgag
gcattgccga agaaaagatt aatggatttg aacacacagc aacagaaact acatgaagtg
                                                                              120
aaacacagga aaaaaaagat aaagaaacga aaagaaaagg gcatcagtga gcttcagcag
                                                                              180
                                                                              240
aagttccatc ggccttacat atgtgtaagc agaggccctg taggagcaga ggcaggggga
aaatacttta agaaataatg tciaaaagtt tttcaaatat gaggaaaaac ataaaaccac agatccaaga agctcaacaa aacaaagcac aagaaacagg aagaaattaa aagttatatc
                                                                              300
                                                                              360
acagtcaaat tgctgaaaac cagcaacaaa gagaatatct taagagtatc agaggaaaag
agattaatga caggccaaga aacaatgaaa acaatacaga tttcttgtag gaaacacaag
acaaaagaca ttttttaaaa ccaaaaggaa aaaaaatgct acattaaaat gtttttacc
                                                                              540
cactgaaagt atatttcaaa acatatttta ggccaggctt ggtggctcac acctgtaatc
                                                                              600
ccagcacttt gggaggccaa ggtgggtgga tcgcttaagg tcaggagttc gagaccagcc
                                                                              660
                                                                              720
tggccaatat agcgaaaccc catctgtact aaaaacacaa aaattagctg ggtgtggtga
                                                                              780
cacatgcctg taatcccagg tactcaggag gctaaggcag gagaattgct tgaactggga
                                                                              840
ggcagaggtg gtgagccaag attgcaccag tgcactccag ccttggtgac agagtgaaac
tccatctcaa aaacaaacaa acaaaataca tatacataaa tatatatgca catatatata
                                                                              900
catatataaa tatatataca catatataaa totatataca tatatacata tatacacata
tataaatcta tatacatata tatacatata taatatattt acatatataa atatatacat
                                                                            1020
atataaatat acatatataa atacatatat aaatatacat atataaatat acatatataa
                                                                            1080
atatacatat ataaatatat acatatataa atatacatat ataaatatat atacatatat
                                                                            1140
aaatatataa atatacaagt atatacaaat atatacatat ataaatgtat atacgtatat
                                                                            1200
                                                                            1260
acatatatat ataaatatat aaaaaaactt ttggctgggc acctttccaa atctcatggc
                                                                            1320
acatataagt ctcatggtaa cctcaaataa aaaaacatat aacagataca ccaaaaaataa
aaaccaataa attaaatcat gccaccagaa gaaattacct tcactaaaag gaacacagga
                                                                             1380
aggaaagaaa gaaggaagag aagaccatga aacaaccaga aaacaaacaa caaaacagca
                                                                            1440
ggagtaattc ctgacttatc aataataatg ctgggtgtaa atggactaaa ctctccaatc
                                                                            1500
aaaagacata gagtggctga atggacgaaa aaaacaagac tcaataatct gttgcctaca
                                                                             1560
agaatatact tcacctataa agggacacat agactgaaaa taaaaggaag gaaaaatatt
                                                                             1620
ctatgcaaat ggaaaccaaa aaaagaacag aactagctac acttatatca gacaaaatag
                                                                            1680
atttcaagac aaaaagtaca aaaagagaca aagtaattat ataataataa agcaaaaaga
                                                                            1740
tataacaatt gtgaatttat atgcgcccaa cactgggaca cccagatata tacagcaaat
                                                                             1800
                                                                             1860
attattagaa ctaaggagag agagagatcc ccatacaata atagctggag acttcacccc
                                                                             1920
gcttttagca ttggacagat catccagaca gaaaatcaac caaaaaattg gacttaatct ataatataga acaaatgtac ctaattgatg tttacaagac atttcatcca gtagttgcag
                                                                             1980
aatatgcatt tittecteag catatggate atteteaagg atagaceata tattaggees
cagaacaage caftaaaaat teaaaaaaat tgageeagge atgatggett atgettgtaa
                                                                             2040
                                                                             2100
ttacagcact ttggggaggg tgaggtggga ggatgtcttg agtacaggag tttgagacca
                                                                             2160
gcctgggcaa aatagtgaga ccctgtctct acaaactttt ttttttaatt agccaggcat
                                                                             2220
agtggtgtgt gcctgtagtc ccagctactt aggaggctga agtgggagga tcacttgagc
                                                                             2280
ccaagagttc aaggctacgg tgagccatga ttgcaacacc acacaccagc cttggtgaca
                                                                             2340
                                                                             2400
gaatgagacc ctgtctcaaa aaaaaaaaaa aaaattgaaa taatataaag catcttctct
ggccacagtg gaacaaaacc agaaatcaac aacaagagga attttgaaaa ctatacaaac
acatgaaaat taaacaatat acttctgaat aaccagtgag tcaatgaaga aattaaaaag
                                                                             2460
                                                                             2520
gaaattgaaa aatttattta agcaaatgat aacggaaaca taacctctca aaacccacgg
                                                                             2580
tatacagcaa aagcagtgct aagaaggaag tttatagcta taagcagcta catcaaaaaa
                                                                             2640
gtagaaaagc caggcgcagt ggctcatgcc tgtaatccca gcactttggg aggccaaggc
gggcagatcg cctgaggtca ggagttcgag accagcctga ccaacacaga gaaaccttgt
                                                                             2760
cgctactaaa aatacaaaat tagctgggca tggtggcaca tgcctgtaat cccagctact
                                                                             2820
cgggaggctg aggcaggata accgcttgaa cccaggaggt ggaggttgcg gtgagccggg
                                                                             2880
                                                                             2940
attgcgccat tggactccag cctgggtaac aagagtgaaa ccctgtctca agaaaaaaaa
aaaagtagaa aaacttaaaa atacaaccta atgatgcacc ttaaagaact agaaaagcaa gagcaaacta aacctaaaat tggtaaaaga aaagaaataa taaagatcag agcagaaata
                                                                             3000
                                                                             3060
aatgaaactg aaagataaca atacaaaaga tcaacaaaat taaaagttgg ttttttgaaa
                                                                             3120
agataaacaa aattgacaaa cctttgccca gactaagaaa aaaggaaaga agacctaaat
```

| aaataaagtc | agagatgaaa | aaaqaqacat | tacaactgat | accacagaaa | ttcaaaggat | 3240 |
|------------|--------------------------|------------|------------|-------------|------------|--------------|
| | tactatgage | | | | | 3300 |
| | | | | | | |
| | gatgcataca | | | | | 3360 |
| cagaccaata | acaataatgg | gattaaagcc | ataataaaaa | atctcctaac | aaagagaagc | 3420 |
| ccannaccca | atggcttccc | toctopattt | taccaatcat | ttaaanaana | atnaattera | 3480 |
| atestastca | anggottete | 22222222 | ******** | ******* | ****** | |
| | aactattctg | | | | | 3540 |
| gccagtatta | ccctgattcc | aaaaccagac | aaaaacacat | caaaaacaaa | caaacaaaaa | 3600 |
| aacagaaaga | aagaaaacta | caggccaata | tecetoatoa | atactoatac | aaaaatcctc | 3660 |
| | | | | | | |
| | tagcaaacca | | | | | 3720 |
| gtgggattta | ttccagggat | ggaaggatgg | ttcaacatat | gcaaatcaat | caatgtgata | 3780 |
| catcatccca | acaaaatgaa | otacaaaaac | tatatoatta | tttcacttta | tocagaaaaa | 3840 |
| | | | | | | |
| | aaattctgca | | | | | 3900 |
| aaacatacag | gccaggcaca | gtggctcaca | cctgcgatcc | cagcactctg | ggaggccaag | 3960 |
| gtgggatgat | tgcttgggcc | caggagtttg | agactagcct | gggcaacaaa | atgagacctg | 4020 |
| | aactttttta | | | | | 4080 |
| | | | | | | |
| | aggctgaggt | | | | | 4140 |
| gccatgaaca | tgtcactgta | ctccagccta | gacaacagaa | caagacccca | ctgaataaga | 4200 |
| | ggagaaggga | | | | | 4260 |
| | ggaaggggaa | | | | | 4320 |
| | | | | | | |
| taataaaagc | cctatatgac | agaccgaggt | agtattatga | ggaaaaactg | aaagcctttc | 4380 |
| ctctaagatc | tggaaaatga | caagggccca | ctttcaccac | totoattcaa | catagtacta | 4440 |
| | | | | | | 4500 |
| | ctagagcaat | | | | | |
| | aattatcctg | | | | | 4560 |
| acaccactaa | aaaactatta | qaqctqaaat | ttaatacaac | aggatacaaa | atcaatotac | 4620 |
| | agtatttcta | | | | | 4680 |
| | | | | | | |
| | taaaattaaa | | | | | 4740 |
| gaaaactata | aaatattgat | aaaagaaatt | gaagagggca | caaaaaaaga | aaagatattc | 4800 |
| | gattggaaga | | | | | 4860 |
| | | | | | | 4920 |
| | tgcaatccct | | | | | |
| caattctaag | atttgtacag | aaccacaaaa | gacccagaat | agccaaagct | atcctgacca | 4980 |
| aaaadaacaa | aactggaagc | atcacattac | ctgacttcaa | attatactac | aaaoctatao | 5040 |
| | tacatggtac | | | | | |
| | | | | | | 5100 |
| tagagaatcc | agaaacaaat | ccatgcatct | acagtgaact | catttttgac | aaaggtgcca | 5160 |
| agaacatact | ttggggaaaa | gataatctct | tcaataaato | atactagaga | aactogatat | 5220 |
| | aataacaata | | | | | 5280 |
| | | | | | | |
| | ggcttaaatc | | | | | 5340 |
| agaaactctc | caggacattg | gagtgggcaa | agacttcttg | agtaattccc | tgcaggcaca | 5400 |
| | gcaaaaacag | | | | | 5460 |
| | | | | | | |
| | atcaacaaag | | | | | 5520 |
| ctattcatct | aacaaggaat | taataaccag | tatatataag | gagctcaaac | tactctataa | 5580 |
| gaaaaacacc | taataagctg | attttcaaaa | ataaqcaaaa | gatctgggta | gacatttctc | 5640 |
| | atacaaatgg | | | | | 5700 |
| | | | | | | |
| | caaatcaaaa | | | | | 5760 |
| tattcaaaag | acaggcaata | acaaatgcca | gtgaggatgt | ggataaaagg | aaacccttgg | 5820 |
| | tgggaatgga | | | | | 5880 |
| | | | | | | |
| | ataaagctac | | | | | 5940 |
| ggaatcagtg | tatcaacaag | ctatctccac | tcccacattt | actgcagcac | tgttcatagc | 6000 |
| agccaaggtt | tggaagcaac | ctcagtgtcc | atcaacagac | gaatggaaaa | agaaaatgtg | 6060 |
| | acaatggagt | | | | | 6120 |
| | | | | | | |
| | ggggcactgg | | | | | 6180 |
| acttttcatg | ttctccctta | cttgtgggag | caaaaattaa | aacaattgac | atagaaatag | 6240 |
| | tggttctaga | | | | | 6300 |
| | | | | | | 6360 |
| | aaaataacta | | | | | |
| aaatgcttga | aggtgacaga | taccccattt | accctgatgt | gattattaca | cattgtatgc | 6420 |
| ctgtatcaaa | atatctcatg | tatgctatag | atataaaccc | tactatatta | aaaattaaaa | 6480 |
| | caggcacggt | | | | | 6540 |
| | | | | | | |
| | ctgaggtcag | | | | | 6600 |
| cctactaaag | atacaaaaat | cagccaggcg | cggtggcaca | tacctgtagt | cccaactact | 6660 |
| caggaggctg | agacaggaga | attgcttgaa | cctgggaggc | ggaggttgca | gtgagccgag | 6720 |
| | tgcactgcag | | | | | 6780 |
| | gattaaaatt | | | | | 6840 |
| | | | | | | |
| | aaattaaaac | | | | | 6900 |
| caatgtatgt | ggggtttcta | gcttctgaag | aagtaaaagt | tatggccacq | atggcagaaa | 6960 |
| | gaacagtgga | | | | | 7020 |
| | | | | | | |
| | caaagacagg | | | | | 7080 |
| actgctaata | atggtgaaag | gtaatctcta | ttaattacca | ataattacag | atatctctaa | 7140 |
| | cagaattggc | | | | | 7200 |
| tettatatac | ttggagattt | tenattatat | attentatt | anttonactt | 22505050 | |
| | | | | | | 7260 |
| | cgaaaaatgg | | | | | 7320 |
| gcaggtggct | ctgtggacct | gagccacttc | aatcttcaao | ggtctctqqc | caagacccao | 7380 |
| | gaggcctgat | | | | | 7440 |
| | | | | | | |
| | tcgttggtga | | | | | 7500 |
| gctctggata | ccatctggaa | aaggcggcca | gcgggaatgc | aaggagtcag | aagcctcctg | 7560 |
| | ggccagcagc | | | | | 7620 |
| | gaagtatggc | | | | | 7680 |
| | | | | | | |
| | gcttgagtta | | | | | 7740 |
| tcaagggaaa | accagacgcc | cgctctgcgg | tcatttacct | ctttcctctc | tccctctctt | 7800 |
| | tttctgatcg | | | | | 7860 |
| | | | | | | |
| yayyactete | ttgcaaaggg | Liccacagac | receipered | yayayaggag | cctgagcctg | 7920 |
| gcttaataac | aaactgggat | gtggctgggg | gcggacagcg | acggcgggat | tcaaagactt | 7980 |
| aattccatga | qtaaattcaa | cctttccaca | tccgaatgaa | tttggatttt | atcttaatat | 8040 |
| | | | | | | -0.0 |
| tttcttaaar | ttcatcaaa+ | aacattcann | agtgcagaaa | tecaaaaaaca | taaaacacca | 2100 |
| actoractet | ttcatcaaat gtttgccaag | aacattcagg | agtgcagaaa | tccaaaggcg | taaaacagga | 8100 8160 |

| cctaagtact | ttttattggt | tttcataagg | tggcttaggg | tgcaagggaa | agtacacgag | 8220 |
|--|--|---|--|--|---------------------------------------|----------------|
| gagaggactg | ggcggcaggg | ctatgagcac | ggcaaggcca | ccggggagag | agtccccggc | 8280 |
| ctgggaggct | gacagcagga | ccactgaccg | tcctccctgg | gagctgccac | attgggcaac | 8340 |
| | ccacgctgcg | | | | | 8400 |
| | ggaagtcacg | | | | | 8460 |
| | gggtgggtca | | | | | 8520 |
| | gctgatgggg | | | | | 8580 |
| tgaatgattc | cagcaacttc | ttcgggtgtg | acaagccatg | acaacactca | gtacaaacac | 8640 |
| cactctttta | ctaggcccac | agagcacggc | ccacacccct | gatatattaa | gagtccagga | 8700 |
| | tgctttcagc | | | | | 8760 |
| | gtagaccctg | | | | | 8820 |
| | catctgccct | | | | | 8880 8940 |
| | tccgcctcca | | | | | 9000 |
| | gcgctactgt tcctcacatg | | | | | 9060 |
| | ctgcgcctcc | | | | | 9120 |
| | tcctggcgtc | | | | | 9180 |
| | gccggggcct | | | | | 9240 |
| | gtctctgtcc | | | | | 9300 |
| | gggcgctctt | | | | | 9360 |
| | gtccacgggc | | | | | 9420 |
| | cacttttctg | | | | | 9480 |
| | cctcttccca | | | | | 9540 |
| | tcacgtgact | | | | | 9600 |
| gctgttttat | tttaatagct | acaaagcagg | gaaatccctg | ctaaaatgtc | ctttaacaaa | 9660 |
| | aaacgggtcc | | | | | 9720 |
| gccgtttata | aagcctgcag | gcatctcaag | ggaattacgc | tgagtcaaaa | ctgccacctc | 9780 |
| | gtacgcaaca | | | | | 9840 |
| gttggggggt | taaggacggt | gggggcagca | gctgggggct | actgcacgca | ccttttacta | 9900 |
| | cctggttctg | | | | | 9960 |
| | ggaacccgga | | | | | 10020 |
| ataatgctct | agagatgccc | acgtcctgat | tcccccaaac | ctgtggacag | aacccgcccg | 10080 |
| | ctttgcaggt | | | | | 10140 10200 |
| | ggcccgaaaa | | | | | 10260 |
| | gcctcaggac | | | | | 10320 |
| | aggcctgcaa cagcaggaag | | | | | 10320 |
| | gtgccatagg | | | | | 10440 |
| atttncanaa | gcaacaggaa | acceatorac | totoaatcta | quattatttc | 22322C222GG | 10500 |
| tttacagaaa | catccaagga | cagggctgaa | atacetecaa | gcaagggcag | oocaoocaco | 10560 |
| agtgatttta | tttagctatt | ttattttatt | tacttacttt | ctgagacaga | attatactct | 10620 |
| | gctggagtgc | | | | | 10680 |
| | ttctcgtgcc | | | | | 10740 |
| | aattttgtat | | | | | 10800 |
| | ctgacctcag | | | | | 10860 |
| | cactgcacct | | | | | 10920 |
| | gtaaggagtt | | | | | 10980 |
| | tctgtaattc | | | | | 11040 |
| | ccacctgtta | | | | | 11100 |
| | ccagtggggt | | | | | 11160 |
| | actgtcctga | | | | | 11220 11280 |
| | ctcctactct | | | | | 11340 |
| | ctcactcctg | | | | | 11400 |
| | ctgtttcatt | | | | | 11460 |
| | ggctggaggg agtgattctc | | | | | 11520 |
| | cagctaattt | | | | | 11580 |
| | | | | | cctctgcctc | |
| ctaaaatact | gggattacag | atataaacca | ccatacccaa | ctcagaattt | actctqttta | 11700 |
| | ggtctgaggt | | | | | 11760 |
| | tttttttatt | | | | | 11820 |
| agacatcatc | agcttttcaa | agacacacta | actgcaccca | taatactggg | gtgtcttctg | 11880 |
| ggtatcagcg | atcttcattg | aatgccggga | ggcgtttcct | cgccatgcac | atggtgttaa | 11940 |
| | ataatcttct | | | | | 12000 |
| | ttctctgcag | | | | | 12060 |
| | ccgccccttt | | | | | 12120 |
| | ggatcactaa | | | | | 12180 |
| | ctaacctcca | | | | | 12240 |
| | cataaataaa | | | | | 12300 12360 |
| | gcgtttgtta gagacccaga | | | | | 12360 |
| | tgggaagtcc | | | | | 12420 |
| | gccgtgtggc | | | | | 12540 |
| | tggagccagg | | | | | 12600 |
| | gatgttggcc | | | | | 12660 |
| | gtgtgaggcg | | | | | 12720 |
| | gacgggaccg | | | | | 12780 |
| | | | | | gtgtcaagga | 12840 |
| | ccicgccgcc | cgagaacttg | | | | |
| gcccaagtcg | cggggaagtg | | | aggtcccgcg | tgcccgtcca | 12900 |
| gggagcaatg | cggggaagtg cgtcctcggg | ttgcagggag ttcgtcccca | gcactccggg gccgcgtcta | cgcgcctccg | tcctcccctt | 12960 |
| gggagcaatg cacgtccggc | cggggaagtg cgtcctcggg attcgtggtg | ttgcagggag ttcgtccca cccggagccc | gcactccggg gccgcgtcta gacgccccgc | cgcgcctccg gtccggacct | tcctccctt ggaggcagcc | 12960 13020 |
| gggagcaatg cacgtccggc ctgggtctcc | cggggaagtg cgtcctcggg | ttgcagggag ttcgtccca cccggagccc agcggccaaa | gcactccggg gccgcgtcta gacgccccgc gggtcgccgc | cgcgcctccg gtccggacct acgcacctgt | tcctccctt ggaggcagcc tcccagggcc | 12960 |

```
gctggggccc tcgctggcgt ccctgcaccc tgggagcgcg agcggcgcg gggcggggaa 13200
gcgcggccca gaccccggg tccgcccgga gcagctgcgc tgtcggggcc aggccgggct 13260
cccagtggat tcgcgggaca agacgcccag gaccgcgctt cccacgtggc ggagggactg 13320
gggacccggg cacccgtct gcccttcac cttccagctc cgcctcctcc gcgcgaccc 13380
cgccccgtcc cgacccctcc cgggtccccg gcccagccc ctccgggccc tcccagcccc tccccttcct ttccgcggcc ccgccctctc ctcgcggcg gagtttcagg cagcgtgcg tcctgctgcg cacgtgggaa gccctggccc cggccacccc cgcgatgccg cgcgctcccc
                                                                                                                          13440
                                                                                                                          13500
gctgccgagc cgtgcgctcc ctgctgcgca gccactaccg cgaggtgctg ccgctggcca
                                                                                                                          13620
cgttcgtgcg gcgcctgggg ccccagggct ggcggctggt gcagcgcggg gacccggcgg
                                                                                                                          13680
ctttccgcgc gctggtggcc cagtgcctgg tgtgcgtgcc ctgggacgca cggccgcccc
                                                                                                                          13740
 ccgccgcccc ctccttccgc caggtgggcc tccccggggt cggcgtccgg ctggggttga
                                                                                                                          13800
gggcggccgg ggggaaccag cgacatgcgg agagcagcgc aggcgactca gggcgcttcc
                                                                                                                          13860
cccgcaggtg tcctgcctga aggagctggt ggcccgagtg ctgcagaggc tgtgcgagcg
                                                                                                                          13920
cggcgcgaag aacgtgctgg cettcggctt cgcgctgctg gacggggcc gcgagggccc ccccgagggc ttcaccacca gcgtgcgcag ctacctgccc aacacggtga cgacgcact gcgggggagc ggggcgtggg ggctgctgct gcgccgcgtg ggcgacgacg tgctggttca cctgctggca cgctgcgcc tctttgtgct ggtggctcc agctgcgcct accaggtgt cgggccgccg ctgtaccagc tcggcgctgc cactcaggcc cggcccccgc cacacgctag
                                                                                                                          13980
                                                                                                                          14040
                                                                                                                          14100
                                                                                                                          14160
                                                                                                                          14220
 tggaccccga aggcgtctgg gatgcgaacg ggcctggaac catagcgtca gggaggccgg
                                                                                                                          14280
 ggtccccctg ggcctgccag ccccgggtgc gaggaggcgc gggggcagtg ccagccgaag
                                                                                                                          14340
 tctgccgttg cccaagaggc ccaggcgtgg cgctgcccct gagccggagc ggacgcccgt
                                                                                                                          14400
 tgggcagggg tcctgggccc acccgggcag gacgcgtgga ccgagtgacc gtggtttctg
                                                                                                                          14460
                                                                                                                          14520
 tgtggtgtca cctgccagac ccgccgaaga agccacctct ttggagggtg cgctctctgg
cacgcgccac teccacecat ecgtgggeeg ceageaceae gegggeeee catecacate geggeeacea egteeetggg acaegeettg tecceeggtg tacgcegaga ecaageaett ectetactee teaggegaea aggageaget geggeeetee tecetactea geteetgag
                                                                                                                          14580
                                                                                                                          14640
                                                                                                                          14700
gatgccagg actggcgctc ggaggctcgt ggagaccatc tttctggggtt ccaggcctg
gatgccaggg actccccgca ggttgccccg cctgccccag cgctactggc aaatgcggcc
                                                                                                                          14760
 cctgtttctg gagctgcttg ggaaccacgc gcagtgcccc tacggggtgc tcctcaagac
                                                                                                                          14880
 gcactgcccg ctgcgagctg cggtcacccc agcagccggt gtctgtgccc gggagaagcc
                                                                                                                          14940
 ccagggetet gtggcggccc ccgaggagga ggacacagac ccccgtcgcc tggtgcagct
                                                                                                                          15000
 gctccgccag cacagcagcc cctggcaggt gtacggcttc gtgcgggcct gcctgcgccg
                                                                                                                          15060
 gctggtgccc ccaggcctct ggggctccag gcacaacgaa cgccgcttcc tcaggaacac
                                                                                                                          15120
caagaagitt atticcctgg ggaagcatgc caagctctcg ctgcaggagc tgacgtggacag
gatgagcgtg cgggactgcg cttggctgcg caaggagccca ggtgaaggagg tggtggccgt
cgagggccca ggccccagag ctgaatgcag taggggctca gaaaaggggg caggcagagc
cctggtcctc ctgtctccat cgtcacgtgg gcacacgtgg cttttcgctc aggacgtcga
gtggacacgg tgatcgagtc gactcccttt agtgagggtt aattgagctc gcggccgc
                                                                                                                          15180
                                                                                                                          15240
                                                                                                                          15300
                                                                                                                          15360
 <210>
 <21.1> 1062
 <212>
             DNA
            Homo sapiens
 <213>
 <221>
             CDS
 <222>
             (1)..(1062)
 <223>
 <400> 2
atg gcc gag gtg ttg cgg acg ctg gcc gga aaa cca aaa tgc cac gca
Met Ala Glu Val Leu Arg Thr Leu Ala Gly Lys Pro Lys Cys His Ala
1 5 10 15
ctt cga cct atg atc ctt ttc cta ata atg ctt gtc ttg ttt ttg ttt
                                                                                                                                48
 Leu Arg Pro Met Ile Leu Phe Leu Ile Met Leu Val Leu Val Leu Phe
20 25 30
 ggt tac ggg gtc cta agc ccc aga agt cta atg cca gga agc ctg gaa
Gly Tyr Gly Val Leu Ser Pro Arg Ser Leu Met Pro Gly Ser Leu Glu
35 40 ... 45
                                                                                                                               144
 cgg ggg ttc tgc atg gct gtt agg gaa cct gac cat ctg cag cgc gtc
Arg Gly Phe Cys Met Ala Val Arg Glu Pro Asp His Leu Gln Arg Val
50 55 60
                                                                                                                               192
 tcg ttg cca agg atg gtc tac ccc cag cca aag gtg ctg aca ccg tgg
Ser Leu Pro Arg Met Val Tyr Pro Gln Pro Lys Val Leu Thr Pro Trp
                                                                                                                               240
aag gat gtc ctc gtg gtg acc cct tgg ctg gct ccc att gtc tgg gag Lys Asp Val Leu Val Thr Pro Trp Leu Ala Pro Ile Val Trp Glu 85 90 95
                                                                                                                               288
ggc aca ttc aac atc gac atc ctc aac gag cag ttc agg ctc cag aac Gly Thr Phe Asn Ile Asp Ile Leu Asn Glu Gln Phe Arg Leu Gln Asn 100 105 110

acc acc att ggg tta act gtg ttt gcc atc aag aaa tac gtg gct ttc Thr Thr Ile Gly Leu Thr Val Phe Ala Ile Lys Lys Tyr Val Ala Phe 115 120 125
                                                                                                                               336
 ctg aag ctg ttc ctg gag acg gcg gag aag cac ttc atg gtg ggc cac
Leu Lys Leu Phe Leu Glu Thr Ala Glu Lys His Phe Met Val Gly His
130 135 140
                                                                                                                               432
 cgt gtc cac tac tat gtc ttc acc gac cag ctg gcc gcg gtg ccc cgc
Arg Val His Tyr Tyr Val Phe Thr Asp Gln Leu Ala Ala Val Pro Arg
145 150 160
                                                                                                                               480
 gtg acg ctg ggg acc ggt cgg cag ctg tca gtg ctg gag gtg cgc gcc
Val Thr Leu Gly Thr Gly Arg Gln Leu Ser Val Leu Glu Val Arg Ala
                                                                                                                               528
```

```
170
576
                                                                                                                                                                   624
 act ccg ctg ttc ggc acc ctg cac ccc ggc ttc tac gga agc agc cgg
Thr Pro Leu Phe Gly Thr Leu His Pro Gly Phe Tyr Gly Ser Ser Arg
225 230 230 240
                                                                                                                                                                    720
 gag gcc ttc acc tac gag cgc cgg ccc cag tcc cag gcc tac atc ccc
Glu Ala Phe Thr Tyr Glu Arg Arg Pro Gln Ser Gln Ala Tyr Ile Pro
245 250 255
                                                                                                                                                                    768
 aag gac gag ggc gat ttc tac tac ctg ggg ggg ttc ttc ggg ggg tcg
Lys Asp Glu Gly Asp Phe Tyr Tyr Leu Gly Gly Phe Phe Gly Gly Ser
260 265 270
gtg caa gag gtg cag cgg ctc acc agg gcc tgc cac cag gcc atg atg val Gln Glu Val Gln Arg Leu Thr Arg Ala Cys His Gln Ala Met Met 275
gtc gac cag gcc aac ggc atc gag gcc gtg tgg cac gac gag agc cac Val Asp Gln Ala Asn Gly Ile Glu Ala Val Trp His Asp Glu Ser His 290
ctg aac aag tac ctg ctg cgc cac aaa ccc acc aag gtg ctc tcc ccc Leu Asn Lys Tyr Leu Leu Arg His Lys Pro Thr Lys Val Leu Ser Pro 305
310
320
320
                                                                                                                                                                    864
                                                                                                                                                                    912
 gag tac ttg tgg gac cag cag ctg ctg ggc tgg ccc gcc gtc ctg agg
Glu Tyr Leu Trp Asp Gln Gln Leu Leu Gly Trp Pro Ala Val Leu Arg
325 330 335
                                                                                                                                                                  1008
 aag ctg agg ttc act gcg gtg ccc aag aac cac cag gcg gtc cgg aac
Lys Leu Arg Phe Thr Ala Val Pro Lys Asn His Gln Ala Val Arg Asn
340 345 350
                                                                                                                                                                  1056
                                                                                                                                                                  1062
  ccg tga
  Pro
```

<210> 3 <211> 353 <212> PRT <213> Homo sapiens

<400> 3

Tyr Lys Arg Trp Gln Asp Val Ser Met Arg Arg Met Glu Met Ile Ser 180 185 190

Asp Phe Cys Glu Arg Arg Phe Leu Ser Glu Val Asp Tyr Leu Val Cys 200 Val Asp Val Asp Met Glu Phe Arg Asp His Val Gly Val Glu Ile Leu Thr Pro Leu Phe Gly Thr Leu His Pro Gly Phe Tyr Gly Ser Ser Arg 225 230 235 240 Glu Ala Phe Thr Tyr Glu Arg Arg Pro Gln Ser Gln Ala Tyr Ile Pro 245 250 255 Lys Asp Glu Gly Asp Phe Tyr Tyr Leu Gly Gly Phe Phe Gly Gly Ser 260 265 270 Val Gln Glu Val Gln Arg Leu Thr Arg Ala Cys His Gln Ala Met Met Val Asp Gln Ala Asn Gly Ile Glu Ala Val Trp His Asp Glu Ser His Leu Asn Lys Tyr Leu Leu Arg His Lys Pro Thr Lys Val Leu Ser Pro 305 310 315 320 Glu Tyr Leu Trp Asp Gln Gln Leu Leu Gly Trp Pro Ala Val Leu Arg 325 330 335 Lys Leu Arg Phe Thr Ala Val Pro Lys Asn His Gln Ala Val Arg Asn 340 345 350 <210> 1065 <211> DNA <212> Homo sapiens <213> <220> <221> CDS <222> (1)..(1065) <223> <400> 4 atg gcc gag gtg ttg cgg acg ctg gcc gga aaa cca aaa tgc cac gca Met Ala Glu Val Leu Arg Thr Leu Ala Gly Lys Pro Lys Cys His Ala 1 5 10 15 10 15
ctt cga cct atg atc ctt ttc cta ata atg ctt gtc ttg gtc ttg
Leu Arg Pro Met Ile Leu Phe Leu Ile Met Leu Val Leu Val Leu Phe
20 25 30
ggt tac ggg gtc cta agc ccc aga agt cta atg cca gga agc ctg gaa
Gly Tyr Gly Val Leu Ser Pro Arg Ser Leu Met Pro Gly Ser Leu Glu
35 40
cgg ggg ttc tgc atg gct gtt agg gaa cct gac cat ctg cag cgc gtc
Arg Gly Phe Cys Met Ala Val Arg Glu Pro Asp His Leu Gln Arg Val
50 55 60 96 144 tcg ttg cca agg atg gtc tac ccc cag cca aag gtg ctg aca ccg tgt Ser Leu Pro Arg Met Val Tyr Pro Gln Pro Lys Val Leu Thr Pro Cys 65 70 75 80 agg aag gat gtc ctc gtg gtg acc cct tgg ctg gct ccc att gtc tgg Arg Lys Asp Val Leu Val Val Thr Pro Trp Leu Ala Pro Ile Val Trp 85 90 95 288 85 90 95 95

gag ggc acg ttc aac atc gac atc ctc aac gag cag ttc agg ctc cag
Clu Gly Thr Phe Asn Ile Asp Ile Leu Asn Glu Gln Phe Arg Leu Gln
100 105 110

aac acc acc att ggg tta act gtg ttt gcc atc aag aaa tac gtg gct
Asn Thr Thr Ile Gly Leu Thr Val Phe Ala Ile Lys Lys Tyr Val Ala
115

ttc ctg aag ctg ttc ctg gag acg gcg gag aag cac ttc atg gtg ggc
Phe Leu Lys Leu Phe Leu Glu Thr Ala Glu Lys His Phe Met Val Gly
130 135

cac cgt gtc cac tac tat gtc ttc acc gac cag ccg gcg gtg ccc
His Arg Val His Tyr Tyr Val Phe Thr Asp Gln Pro Ala Ala Val Pro
145

cgc gtg acg ctg ggg acc ggt cgg cag ctg tca gtg ctg ggg
Arg Val Thr Leu Gly Thr Gly Arg Gln Leu Ser Val Leu Glu Val Gly
165

gcc tac aag ccg tgg cag gac gtc cat gag atg atg atc 384 432 528 gcc tac aag cgc tgg cag gac gtg tcc atg cgc cgc atg gag atg atc Ala Tyr Lys Arg Trp Gln Asp Val Ser Met Arg Arg Met Glu Met Ile 180 185 190 576 agt gac ttc tgc gag cgg cgc ttc ctc agc gag gtg gat tac ctg gtg Ser Asp Phe Cys Glu Arg Arg Phe Leu Ser Glu Val Asp Tyr Leu Val

<210> \$ <211> 354 <212> PRT <213> Homo sapiens

<400> 5

 Met 1
 Ala
 Glu
 Val
 Leu
 Arg
 Thr
 Leu
 Ala
 Gly
 Lys
 Pro
 Cys
 His
 Ala

 Leu
 Arg
 Pro
 Met
 Ile
 Leu
 Phe
 Leu
 Ile
 Met
 Leu
 Val
 Leu
 Phe

 Gly
 Tyr
 Gly
 Val
 Leu
 Ser
 Pro
 Arg
 Ger
 Leu
 Met
 Pro
 Arg
 Gu
 Pro
 Arg
 Fru
 Gu
 Pro
 Arg
 Val
 Arg
 Val
 Arg
 Val
 Pro
 Arg
 Gu
 Pro
 Arg
 Fru
 Pro
 Arg
 Fru
 Fru
 Pro
 Arg
 Fru
 Pro
 Arg
 Fru
 Pro
 Arg
 Fru
 Pro
 Arg
 Fru
 Arg
 Fru
 Pro
 Arg
 Pro
 Arg
 Pro

 Leu Thr
 Pro Leu Phe Cally Thr Leu His Pro Ser 235
 Pro Ser 240

 Arg Glu Ala Phe Thr Tyr Glu Arg Arg Pro 250
 Gln Ser Gln Ala Tyr Ile 255

 Pro Lys Asp Glu Gly Asp Phe Tyr Tyr Met Cally Ala Phe Phe Gly Gly 260
 Glu Gly Asp Phe Tyr Tyr Met Gly Ala Phe Phe Gly Gly 270

 Ser Val Gln Glu Val Gln Arg Leu Thr Arg Ala Cys His Gln Ala Met 275
 Glu Ala Asn Gly Ile Glu Ala Val Trp His Asp Glu Ser 295

 Met Val Asp Gln Ala Asn Gly 295
 Ile Glu Ala Val Trp His Asp Glu Ser 310

 His Leu Asn Lys Tyr Leu Leu Leu Arg His Lys Pro 310
 Thr Lys Val Leu Ser 320

 Pro Glu Tyr Leu Trp Asp Gln Gln Leu Leu Gly Trp Pro Ala Val Arg 335
 Leu Arg Arg Ala Val Arg 345

 Asn Pro
 Asn Pro

<210> 6 <211> 376 <212> PRT <213> Platyrrhinus helleri <400> 6

Met Asn Val Lys Gly Lys Val Ile Leu Ser Met Leu Val Val Ser Thr 1 10 15 Val Ile Val Val Phe Trp Glu Tyr Ile Asn Ser Pro Glu Gly Ser Phe 20 25 30 Leu Trp Ile Tyr His Ser Lys Asn Pro Glu Val Asp Asp Ser Ser Ala 35 40 45 . Gln Lys Asp Trp Trp Phe Pro Gly Trp Phe Asn Asn Gly Ile His Asn 50 55 60 Tyr Gln Gln Glu Glu Glu Asp Thr Asp Lys Glu Lys Gly Arg Glu Glu 65 70 75 80 Glu Gln Lys Lys Glu Asp Asp Thr Thr Glu Leu Arg Leu Trp Asp Trp 85 90 95 Phe Asn Pro Lys Lys Arg Pro Glu Val Met Thr Val Thr Gln Trp Lys 100 105 110 Ala Pro Val Val Trp Glu Gly Thr Tyr Asn Lys Ala Ile Leu Glu Asn 115 120 125 Tyr Tyr Ala Lys Gln Lys Ile Thr Val Gly Leu Thr Val Phe Ala Ile 130 140 Gly Arg Tyr Ile Glu His Tyr Leu Glu Glu Phe Val Thr Ser Ala Asn 145 150 155 160 Arg Tyr Phe Met Val Gly His Lys Val Ile Phe Tyr Val Met Val Asp 165 170 175 Asp Val Ser Lys Ala Pro Phe Ile Glu Leu Gly Pro Leu Arg Ser Phe 180 185 190 Lys Val Phe Glu Val Lys Pro Glu Lys Arg Trp Gln Asp Ile Ser Met 195 200 Met Arg Met Lys Thr Ile Gly Glu His Ile Leu Ala His Ile Gln His 210 220 Glu Val Asp Phe Leu Phe Cys Met Asp Val Asp Gln Val Phe Gln Asp 225 230 240 His Phe Gly Val Glu Thr Leu Gly Gln Ser Val Ala Gln Leu Gln Ala 245 250 255

```
Trp Trp Tyr Lys Ala Asp Pro Asp Asp Phe Thr Tyr Glu Arg Arg Lys 260 265 270
Glu Ser Ala Ala Tyr Ile Pro Phe Gly Gln Gly Asp Phe Tyr Tyr His 275 280 285
Ala Ala Ile Phe Gly Gly Thr Pro Ile Gln Val Leu Asn Ile Thr Gln 290 \phantom{\bigg|}295\phantom{\bigg|}300\phantom{\bigg|}
Glu Cys Phe Lys Gly Ile Leu Leu Asp Lys Lys Asn Asp Ile Glu Ala
305 310 320
Glu Trp His Asp Glu Ser His Leu Asn Lys Tyr Phe Leu Leu Asn Lys 325 330 335
Pro Ser Lys Ile Leu Ser Pro Glu Tyr Cys Trp Asp Tyr His Ile Gly 340 345 350
Leu Pro Ser Asp Ile Lys Thr Val Lys Leu Ser Trp Gln Thr Lys Glu 355 360 365
Tyr Asn Leu Val Arg Lys Asn Val
370 375
<210> 7
<211> 227
<212> PRT
<213> Homo sapiens
Arg Tyr Asn Asp His Tyr Leu Glu Glu Phe Ile Thr Ser Ala Asn Arg 1 \phantom{\bigg|} 5 \phantom{\bigg|} 10 \phantom{\bigg|} 15
Tyr Phe Met Val Gly His Lys Val Ile Phe Tyr Ile Met Val Asp Asp 20 25 30
Val Ser Lys Leu Pro Phe Ile Glu Leu Gly Pro Leu His Ser Phe Lys
35 40 45
Met Phe Glu Val Lys Pro Glu Lys Arg Trp Gln Asp Ile Ser Met Met 50 60
Arg Met Lys Ile Thr Gly Glu His Ile Leu Ala His Ile Gln His Glu 65 70 75 80
Val Asp Phe Leu Phe Cys Met Asp Val Asp Gln Val Phe Gln Asp His 85 90 95
Phe Gly Val Glu Thr Leu Gly Gln Ser Val Ala Gln Leu Gln Trp Arg 100 \hspace{1.5cm} 105 \hspace{1.5cm} 110
Tyr Lys Ala Asp Pro Tyr Asp Phe Thr Glu Arg Trp Lys Glu Ser Ala
115 120 125
Gly Tyr Ile Pro Phe Gly Gly Asp Phe Tyr Tyr His Ala Ala Ile Ser
130 140
Gly Gly Thr Pro Ile Gln Val Leu Asn Ile Thr Gln Glu Cys Phe Lys
145 150 155 160
Gly Ile Leu Leu Asp Lys Lys Asn Asp Ile Glu Ala Lys Trp His Asp 165 170 175
Glu Ser His Leu Asn Lys Tyr Phe Leu Leu Asn Lys Pro Ser Lys Ile 180 $180\,
Leu Ser Leu Lys Tyr Cys Trp Asp Tyr His Ile Gly Leu Pro Ser Asp
195 200 205
Ile Lys Thr Val Lys Ser Trp Gln Thr Lys Glu Tyr Asn Leu Val Arg 210 220
Asn Asn Val
225
<211> 369
<212> PRT
```

9

<213> Ovis aries <400> 8

Met Asn Val Lys Gly Lys Val Ile Leu Ser Met Leu Val Val Ser Thr 1 10 15 Val Ile Val Val Phe Trp Glu Tyr Ile His Ser Pro Glu Gly Ser Leu 20 25 30 Phe Trp Ile Asn Pro Ser Arg Asn Pro Glu Val Ser Gly Gly Ser Ser 40 45Glu Glu Asp Glu Asp Val Asp Glu Glu Lys Glu Gln Arg Lys Glu Asp 65 70 75 80 Lys Ser Lys Leu Lys Leu Ser Asp Trp Phe Asn Pro Phe Lys Arg Pro 85 90 Glu Val Val Thr Met Thr Asp Trp Lys Ala Pro Val Val Trp Glu Gly 100 105 110 Thr Tyr Asn Arg Ala Val Leu Asp Asp Tyr Tyr Ala Lys Gln Lys Ile 115 120 125 Thr Val Gly Leu Thr Val Phe Ala Val Gly Arg Tyr Ile Glu His Tyr 130 140 Leu Glu Glu Phe Leu Thr Ser Ala Asn Lys His Phe Met Val Gly His 145 150150155 Arg Val Ile Phe Tyr Val Met Val Asp Asp Val Ser Arg Met Pro Leu 165 170 175 Ile Glu Leu Gly Pro Leu Arg Ser Phe Lys Val Phe Glu Val Lys Pro $180 \ \ \,$ 180 $\ \ \,$ 185 $\ \ \,$ 190 $\ \ \,$ Glu Arg Arg Trp Gln Asp Val Ser Met Val Arg Met Lys Thr Ile Gly 195 200 205 Glu His Ile Val Ala His Ile Gln Arg Glu Val Asp Phe Leu Phe Cys 210 220 Met Asp Val Asp Gln Val Phe Gln Asp Glu Phe Gly Val Glu Thr Leu 225 230 235 240 Gly Glu Ser Val Ala Gln Leu Gln Ala Trp Trp Tyr Lys Ala Asp Pro 245 250 255 Asp Glu Phe Thr Tyr Glu Arg Arg Lys Glu Ser Ala Ala Tyr Ile Pro 260 265 270 Phe Gly Glu Gly Asp Phe Tyr Tyr His Ala Ala Ile Phe Gly Gly Thr 275 280 285 Pro Thr Gln Val Leu Asn Ile Thr Gln Glu Cys Phe Lys Gly Ile Leu 290 300 Lys Asp Lys Lys Asp Asp Ile Glu Ala Gln Trp His Asp Glu Ser His 305 310 315 Leu Asn Lys Tyr Phe Leu Leu Asn Lys Pro Thr Lys Ile Leu Ser Pro 325 330 335 Glu Tyr Cys Trp Asp Tyr His Ile Gly Leu Pro Ala Asp Ile Lys Leu 340 345 350 Val Lys Met Ser Trp Gln Thr Lys Glu Tyr Asn Val Val Arg Asn Asn 355 360 365

۷a٦

<210>

<211> 368 <212> PRT <213> Bos taurus

<400> 9

Met Asn Val Lys Gly Lys Val Ile Leu Ser Met Leu Val Val Ser Thr 1 10 15 Val Ile Val Val Phe Trp Glu Tyr Ile His Ser Pro Glu Gly Ser Leu 20 25 30 Phe Trp Ile Asn Pro Ser Arg Asn Pro Glu Val Gly Gly Ser Ser Ile 35 40 45 Gln Lys Gly Trp Trp Leu Pro Arg Trp Phe Asn Asn Gly Tyr His Glu 50 60 Glu Asp Gly Asp Ile Asn Glu Glu Lys Glu Gln Arg Asn Glu Asp Glu 65 70 75 80 Ser Lys Leu Lys Leu Ser Asp Trp Phe Asn Pro Phe Lys Arg Pro Glu 85 90 95 Val Val Thr Met Thr Lys Trp Lys Ala Pro Val Val Trp Glu Gly Thr 100 105 110 Tyr Asn Arg Ala Val Leu Asp Asn Tyr Tyr Ala Lys Gln Lys Ile Thr 115 120 125 Val Gly Leu Thr Val Phe Ala Val Gly Arg Tyr Ile Glu His Tyr Leu 130 135 140 Glu Glu Phe Leu Thr Ser Ala Asn Lys His Phe Met Val Gly His Pro 145 150 160 Val Ile Phe Tyr Ile Met Val Asp Asp Val Ser Arg Met Pro Leu Ile $165 \hspace{0.25cm} 170 \hspace{0.25cm} 175$ Glu Leu Gly Pro Leu Arg Ser Phe Lys Val Phe Lys Ile Lys Pro Glu 180 185 190 Lys Arg Trp Gln Asp Ile Ser Met Met Arg Met Lys Thr Ile Gly Glu $195 \hspace{0.5cm} 200 \hspace{0.5cm} 205 \hspace{0.5cm}$ His Ile Val Ala His Ile Gln His Glu Val Asp Phe Leu Phe Cys Met 210 215 220 Asp Val Asp Gln Val Phe Gln Asp Lys Phe Gly Val Glu Thr Leu Gly 225 230 235 240 Glu Ser Val Ala Gln Leu Gln Ala Trp Trp Tyr Lys Ala Asp Pro Asn 245 250 255 Asp Phe Thr Tyr Glu Arg Arg Lys Glu Ser Ala Ala Tyr Ile Pro Phe 260 265 270 Gly Glu Gly Asp Phe Tyr Tyr His Ala Ala Ile Phe Gly Gly Thr Pro 275 280 285 Thr Gln Val Leu Asn Ile Thr Gln Glu Cys Phe Lys Gly Ile Leu Lys 290 295 300 Asp Lys Lys Asn Asp Ile Glu Ala Gln Trp His Asp Glu Ser His Leu 305 310 315 320 Asn Lys Tyr Phe Leu Leu Asn Lys Pro Thr Lys Ile Leu Ser Pro Glu 325 330 335 Tyr Cys Trp Asp Tyr His Ile Gly Leu Pro Ala Asp Ile Lys Leu Val \$340\$Lys Met Ser Trp Gln Thr Lys Glu Tyr Asn Val Val Arg Asn Asn Val 355 360 365

<210> 10 <211> 373

<212> PRT <213> Sus scrofa

<400> 10

Met Asn Val Lys Gly Arg Val Val Leu Ser Met Leu Leu Val Ser Thr 1 $$ 10 $$ 15

```
Phe Trp Ile Tyr Gln Ser Lys Asn Pro Glu Val Gly Ser Ser Ala Gln 35 40
Arg Gly Trp Trp Phe Pro Ser Trp Phe Asn Asn Gly Thr His Ser Tyr 50 60
His Glu Glu Glu Asp Ala Ile Gly Asn Glu Lys Glu Gln Arg Lys Glu
65 70 75 80
Asp Asn Arg Gly Glu Leu Pro Leu Val Asp Trp Phe Asn Pro Glu Lys 85 \\ 90 \cdot \cdot \cdot \cdot 95 \\
Arg Pro Glu Val Val Thr Ile Thr Arg Trp Lys Ala Pro Val Val Trp
100 105 110
Glu Gly Thr Tyr Asn Arg Ala Val Leu Asp Asn Tyr Tyr Ala Lys Gln 115 120 125
Lys Ile Thr Val Gly Leu Thr Val Phe Ala Val Gly Arg Tyr Ile Glu
130 135 140
His Tyr Leu Glu Glu Phe Leu Ile Ser Ala Asn Thr Tyr Phe Met Val
145 150 155 160
Gly His Lys Val Ile Phe Tyr Ile Met Val Asp Asp Ile Ser Arg Met 165 170 170
Pro Leu Ile Glu Leu Gly Pro Leu Arg Ser Phe Lys Val Phe Glu Ile
180 185 190
Lys Ser Glu Lys Arg Trp Gln Asp Ile Ser Met Met Arg Met Lys Thr 195 200 205
Ile Gly Glu His Ile Leu Ala His Ile Gln His Glu Val Asp Phe Leu 210 215 220
Phe Cys Met Asp Val Asp Gln Val Phe Gln Asn Asn Phe Gly Val Glu
225 235 240
Thr Leu Gly Gln Ser Val Ala Gln Leu Gln Ala Trp Trp Tyr Lys Ala 245 250 250
His Pro Asp Glu Phe Thr Tyr Glu Arg Arg Lys Glu Ser Ala Ala Tyr 260 265 270
Ile Pro Phe Gly Gln Gly Asp Phe Tyr Tyr His Ala Ala Ile Phe Gly 275 280 285
Gly Thr Pro Thr Gln Val Leu Asn Ile Thr Gln Glu Cys Phe Lys Gly 290 295 300
Ile Leu Gln Asp Lys Glu Asn Asp Ile Glu Ala Glu Trp His Asp Glu
305 310 315 320
Ser His Leu Asn Lys Tyr Phe Leu Leu Asn Lys Pro Thr Lys Ile Leu
325 330 335
Ser Pro Glu Tyr Cys Trp Asp Tyr His Ile Gly Met Ser Val Asp Ile 340 345 350
Arg Ile Val Lys Ile Ala Trp Gln Lys Lys Glu Tyr Asn Leu Val Arg 355 360 365
Asn Asn Ile
370
```

<210> 11 <211> 359 <212> PRT <213> Mus musculus

<400> 11

Met Asn Val Lys Gly Lys Val Ile Leu Leu Met Leu Ile Val Ser Thr 1 10 15 Val Val Val Val Phe Trp Glu Tyr Val Asn Arg Ile Pro Glu Val Gly

Glu Asn Arg Trp Gln Lys Asp Trp Trp Phe Pro Ser Trp Phe Lys Asn 35 40 45

Gly Thr His Ser Tyr Gln Glu Asp Asn Val Glu Gly Arg Arg Glu Lys
50 55 60

Gly Arg Asn Gly Asp Arg Ile Glu Glu Pro Gln Leu Trp Asp Trp Phe 65 70 75 80

Asn Pro Lys Asn Arg Pro Asp Val Leu Thr Val Thr Pro Trp Lys Ala 85 90 95

Pro Ile Val Trp Glu Gly Thr Tyr Asp Thr Ala Leu Leu Glu Lys Tyr 100 105 110

Tyr Ala Thr Gln Lys Leu Thr Val Gly Leu Thr Val Phe Ala Val Gly 115 120 125

Lys Tyr Ile Glu His Tyr Leu Glu Asp Phe Leu Glu Ser Ala Asp Met 130 140

Tyr Phe Met Val Gly His Arg Val Ile Phe Tyr Val Met Ile Asp Asp 145 150 155 160

Thr Ser Arg Met Pro Val Val His Leu Asn Pro Leu His Ser Leu Gln 165 170 175

Val Phe Glu Ile Arg Ser Glu Lys Arg Trp Gln Asp Ile Ser Met Met 180 185 190

Arg Met Lys Thr Ile Gly Glu His Ile Leu Ala His Ile Gln His Glu 195 200 205

Val Asp Phe Leu Phe Cys Met Asp Val Asp Gln Val Phe Gln Asp Asn 210 215 220

Phe Gly Val Glu Thr Leu Gly Gln Leu Val Ala Gln Leu Gln Ala Trp 225 230 235 240

Trp Tyr Lys Ala Ser Pro Glu Lys Phe Thr Tyr Glu Arg Arg Glu Leu 245 250 . 255

Ser Ala Ala Tyr Ile Pro Phe Gly Glu Gly Asp Phe Tyr Tyr His Ala 260 265 270

Ala Ile Phe Gly Gly Thr Pro Thr His Ile Leu Asn Leu Thr Arg Glu 275 280 285

Cys Phe Lys Gly Ile Leu Gln Asp Lys Lys His Asp Ile Glu Ala Gln 290 295 300

Trp His Asp Glu Ser His Leu Asn Lys Tyr Phe Leu Phe Asn Lys Pro 305 310 310 320

Thr Lys Ile Leu Ser Pro Glu Tyr Cys Trp Asp Tyr Gln Ile Gly Leu 325 330 335

Pro Ser Asp Ile Lys Ser Val Lys Val Ala Trp Gln Thr Lys Glu Tyr 340 345 350

Asn Leu Val Arg Asn Asn Val

376 PRT

Artificial

<400> 12

Met Asn Val Lys Gly Lys Val Ile Leu Ser Met Leu Val Val Ser Thr 1 $$ 5

Val Ile Val Val Phe Trp Glu Tyr Ile Asn Ser Pro Glu Gly Ser Phe 20 25 30

Leu Trp Ile Tyr His Ser Lys Asn Pro Glu Val Asp Asp Ser Ser Ala 35 40 45

Gln Lys Asp Trp Trp Phe Pro Gly Trp Phe Asn Asn Gly Ile His Asn 50 55 60 Tyr Gln Glu Glu Glu Asp Thr Asp Lys Glu Lys Gly Arg Glu Glu 65 70 75 80 Glu Gln Lys Lys Glu Asp Asp Thr Thr Glu Leu Arg Leu Trp Asp Trp 85 90 95 Phe Asn Pro Lys Lys Arg Pro Glu Val Met Thr Val Thr Gln Trp Lys 100 105 110 Ala Pro Val Val Trp Glu Gly Thr Tyr Asn Lys Ala Ile Leu Glu Asn 115 120 125 Tyr Tyr Ala Lys Gln Lys Ile Thr Val Gly Leu Thr Val Phe Ala Ile 130 140 Gly Arg Tyr Ile Glu His Tyr Leu Glu Glu Phe Leu Thr Ser Ala Asn 145 150 155 160 Arg Tyr Phe Met Val Gly His Lys Val Ile Phe Tyr Val Met Val Asp 165 $$ 170 $$ Met Val Asp Asp Val Ser Lys Ala Pro Phe Ile Glu Leu Gly Pro Leu Arg Ser Phe 180 185 190 Lys Val Phe Glu Val Lys Pro Glu Lys Arg Trp Gln Asp Ile Ser Met 195 200 Met Arg Met Lys Thr Ile Gly Glu His Ile Leu Ala His Ile Gln His 210 220 Glu Val Asp Phe Leu Phe Cys Met Asp Val Asp Gln Val Phe Gln Asp 225 230 240 His Phe Gly Val Glu Thr Leu Gly Gln Ser Val Ala Gln Leu Gln Ala 255 Trp Trp Tyr Lys Ala Asp Pro Asp Asp Phe Thr Tyr Glu Arg Arg Lys 260 265 270Glu Ser Ala Ala Tyr Ile Pro Phe Gly Gln Gly Asp Phe Tyr Tyr His 275 280 285 Ala Ala Ile Phe Gly Gly Thr Pro Ile Gln Val Leu Asn Ile Thr Gln 290 295 300 GTu Cys Phe Lys GTy ITe Leu Leu Asp Lys Lys Asn Asp ITe GTu ATa 305 310 315 320 Glu Trp His Asp Glu Ser His Leu Asn Lys Tyr Phe Leu Leu Asn Lys 325 330 335 Pro Ser Lys Ile Leu Ser Pro Glu Tyr Cys Trp Asp Tyr His Ile Gly 340 345 350 Leu Pro Ser Asp Ile Lys Thr Val Lys Leu Ser Trp Gln Thr Lys Glu 355 360 365 Tyr Asn Leu Val Arg Lys Asn Val

<210> 13 <211> 376 <212> PRT <213> Artificial

<400> 13

Tyr Gln Gln Glu Glu Glu Asp Thr Asp Lys Glu Lys Gly Arg Glu Glu 65 70 75 80

```
Glu Gln Lys Lys Glu Asp Asp Thr Thr Glu Leu Arg Leu Trp Asp Trp
85 90 95
Phe Asn Pro Lys Lys Arg Pro Glu Val Met Thr Val Thr Gln Trp Lys 100 105 110
Ala Pro Val Val Trp Glu Gly Thr Tyr Asn Lys Ala Ile Leu Glu Asn
115 120 125
Tyr Tyr Ala Lys Gln Lys Ile Thr Val Gly Leu Thr Val Phe Ala Ile
130 140
Gly Arg Tyr Ile Asp His Tyr Leu Glu Glu Phe Leu Thr Ser Ala Asn 145 150 155
Arg Tyr Phe Met Val Gly His Lys Val Ile Phe Tyr Ile Met Val Asp 165 \hspace{0.25cm} 170 \hspace{0.25cm} 175 \hspace{0.25cm}
Asp Val Ser Lys Ala Pro Phe Ile Glu Leu Gly Pro Leu Arg Ser Phe 180 185 190
Lys Val Phe Glu Val Lys Pro Glu Lys Arg Trp Gln Asp Ile Ser Met 195 200 205
Met Arg Met Lys Ile Thr Gly Glu His Ile Leu Ala His Ile Gln His 210 220
Glu Val Asp Phe Leu Phe Cys Met Asp Val Asp Gln Val Phe Gln Asp 225 230 240
His Phe Gly Val Glu Thr Leu Gly Gln Ser Val Ala Gln Leu Gln Ala 245 250 255
Trp Trp Tyr Lys Ala Asp Pro Asp Asp Phe Thr Tyr Glu Arg Arg Lys 260 265 270
Glu Ser Ala Gly Tyr Ile Pro Phe Gly Gln Gly Asp Phe Tyr Tyr His 275 280 285
Ala Ala Ile Phe Gly Gly Thr Pro Ile Gln Val Leu Asn Ile Thr Gln
290 295 300
Glu Cys Phe Lys Gly Ile Leu Leu Asp Lys Lys Asn Asp Ile Glu Ala
305 310 320
Glu Trp His Asp Glu Ser His Leu Asn Lys Tyr Phe Leu Leu Asn Lys 325 330 335
Pro Ser Lys Ile Leu Ser Pro Glu Tyr Cys Trp Asp Tyr His Ile Gly 340 345 350
Leu Pro Ser Asp Ile Lys Thr Val Lys Leu Ser Trp Gln Thr Lys Glu 355
Tyr Asn Leu Val Arg Lys Asn Val
<210> 14
<211> 1131
<212> DNA
<213>
       Platyrrhinus helleri
<400> 14
atgaatgtca aaggaaaagt aattctgtcg atgctggttg tctcaactgt gattgttgtg
                                                                                       60
ttttgggaat atatcaacag cccagaaggc tctttcttgt ggatatatca ctcaaagaac
                                                                                     120
                                                                                      180
ccagaagttg atgacagcag tgctcagaag gactggtggt ttcctggctg gtttaacaat
gggatccaca attatcaaca agaggaagaa gacacagaca aagaaaaagg aagaggag
gaacaaaaaa aggaagatga cacaacagag cttcggctat gggactggtt taatccaaag
aaacgcccag aggttatgac agtgacccaa tggaaggcgc cggttgtgtg ggaaggcact
                                                                                      240
                                                                                      300
                                                                                      360
tacaacaaag ccatcctaga aaattattat gccaaacaga aaattaccgt ggggttgacg
                                                                                      420
gtttttgcta ttggaagata tattgagcat tacttggagg agttcgtaac atctgctaat
                                                                                      480
aggtactica tggtcggcca caaagtcata ttttatgtca tggtggatga tgtctccaag
                                                                                      540
gcgccgttta tagagctggg tcctctgcgt tccttcaaag tgtttgaggt caagccagag
                                                                                      600
aagaggtggc aagacatcag catgatgcgt atgaagacca tcgggggagca catcttggcc
                                                                                      660
cacatccaac acgaggttga cttcctcttc tgcatggatg tggaccaggt cttccaagac
                                                                                      720
cattttgggg tagagaccct gggccagtcg gtggctcagc tacaggcctg gtggtacaag
gcagatcctg atgactttac ctatgagagg cggaaagagt cggcagcata tattccattt
                                                                                      780
```

```
ggccaggggg attittatta ccatgcagcc attittggag gaacaccgat tcaggttctc
                                                                                        900
aacatcaccc aggagtgctt taagggaatc ctcctggaca agaaaaatga catagaagcc
                                                                                        960
gagtggcatg atgaaagcca cctaaacaag tatttccttc tcaacaaacc ctctaaaatc
ttatctccag aatactgctg ggattatcat ataggcctgc cttcagatat taaaactgtc
                                                                                      1020
                                                                                      1080
aagctatcat ggcaaacaaa agagtataat ttggttagaa agaatgtctg a
                                                                                      1131
<210>
<211>
        755
<212>
        DNA
        Homo sapiens
<213>
<400> 15
cagcttgtgg tttctttcag gaatcccaga ggataaatgt tttgctttc ttctttgttt cagatataat gatcattact tggaggagtt cataacatct gctaataggt acttcatggt
                                                                                        120
tggccacaaa gtcatatttt acatcatggt ggatgatgtc tccaagctgc cgtttataga
gctgggtcct ctgcattcct tcaaaatgtt tgaggtcaag ccagagaaga ggtggcaaga
catcagcatg atgcgtatga agatcactgg ggagcacatc ttggcccaca tccaacacga
                                                                                        300
ggtcgacttc ctcttctgca tggatgtgga ccaggtcttc caagaccatt ttggggtgga
                                                                                        360
gaccetagge cagteagtgg cteagetaca ggetggeggt acaaggeaga tecetatgae
tttacctagg agaggtggaa agagteagea ggatacatte catttggeca ggggattttt
                                                                                        420
                                                                                        480
attaccatgc agccatttct ggaggaacac ccattcaggt tctcaacatc acccaggagt
                                                                                       540
gctttaaggg aatcctcctg gacaagaaaa atgacataga agccaagtgg catgatgaaa
                                                                                       600
gccacctaaa caagtattic cttctcaata aaccctctaa aatcttatcc ctaaaatact
                                                                                        660
gctgggatta tcatataggc ctgccttcag atattaaaac tgtcaagtga tcgtggcaga
                                                                                       720
caaaagagta taatttggtt agaaataatg tctga
<210>
        16
<211>
        1131
<212>
        DNA
        Artificial
<213>
<400> 16
atgaatgtca aaggaaaagt aattctgtcg atgctggttg tctcaactgt gattgttgtg
ttttgggaat atatcaacag cccagaaggc tctttcttgt ggatatatca ctcaaagaac
ccagaagttg atgacagcag tgctcagaag gactggtggt ttcctggctg gtttaacaat
                                                                                       180
gggatccaca attatcaaca agaggaagaa gacacagaca aagaaaaagg aagagaggag
                                                                                       240
gaacaaaaaa aggaagatga cacaacagag cttcggctat gggactggtt taatccaaag
aaacgcccag aggttatgac agtgacccaa tggaaggcgc cggttgtgtg ggaaggcact
                                                                                       300
                                                                                       360
tacaacaaag ccatcctaga aaattattat gccaaacaga aaattaccgt ggggttgacg
                                                                                        420
gtttttgcta ttggaagata tattgatcat tacttggagg agttcttaac atctgctaat
                                                                                        480
aggtactica tggttggcca caaagtcata tittacatca tggtggatga tgtctccaag
                                                                                        540
gcgccgttta tagagctggg tcctctgcgt tccttcaaag tgtttgaggt caagccagag
aagaggtggc aagacatcag catgatgcgt atgaagatca ctggggagca catcttggcc
                                                                                        660
cacatccaac acgaggtcga cttcctcttc tgcatggatg tggaccaggt cttccaaggac cattttgggg tggagaccct aggccagtca gtggctcagc tacaggcctg gtggtacaag gcagatcccg atgactttac ctatgagagg cggaaagagt cagcaggata cattccattt ggccaggggg atttttatta ccatgcagcc atttttggag gaacacccat tcaggttcc
                                                                                       720
                                                                                       780
                                                                                       840
                                                                                       900
aacatcaccc aggagtgctt taagggaatc ctcctggaca agaaaaatga catagaagcc
                                                                                       960
gagtggcatg atgaaagcca cctaaacaag tatttccttc tcaataaacc ctctaaaatc
                                                                                      1020
ttatccccag aatactgctg ggattatcat ataggcctgc cttcagatat taaaactgtc
                                                                                      1080
aagctatcgt ggcagacaaa agagtataat ttggttagaa ataatgtctg a
                                                                                      1131
<211> 1303
<212>
        DNA
<213>
        Ovis aries
<400> 17
agccgaggac gccgccgggg agccgaggct ccggccagcc cccagcgcgc ccagcttctg
cagatcagga gtcagaacgc tgcaccttcg cttcctccca gccctgcctc cttctgcaaa
                                                                                       120
acggagetea atagaacttg gtacttttgc cttttactct gggaggagag aagcagacga
                                                                                       180
tgaggagaaa ataatgaatg tcaaaggaaa agtgattctg tcaatgctgg ttgtctcaac
                                                                                       240
tgtcattgtt gtgttttggg aatatatcca cagcccagaa ggctctttgt tctggataaa
cccatcaaga aacccagaag tcagtggcgg cagcagcatt cagaagggct ggtggtttcc
                                                                                       360
gagatggttt aacaatggtt accaagaaga agatgaagac gtagacgaag aaaaggaaca
                                                                                       420
aagaaaggaa gacaaaagca agcttaagct atcggactgg ttcaacccat ttaaacgccc
tgaggttgtg actatgacag attggaaggc acccgtggtg tgggaaggca cttacaacag
agccgtctta gacgattact acgccaagca gaaaattacc gtcggcctga cggttttcgc
                                                                                       480
                                                                                       540
                                                                                       600
cgtcggaaga tacattgagc attacttgga ggagttctta acgtctgcta ataagcactt
                                                                                       660
catggttggc caccgagtca tcttttacgt catggtggac gacgtctcca ggatgccttt
                                                                                       720
gatagagetg ggccctctgc gctccttcaa agtgtttgag gtcaagcctg agaggaggtg
                                                                                       780
gcaggacgtc agcatggtgc gcatgaagac catcggggag cacatcgtgg cccacatcca
gcgtgaggtt gacttcctct tctgcatgga cgtggaccag gtcttccaag acgagttcgg
                                                                                       900
ggtggagacc ctgggtgagt cggtggccca gctacaggcc tggtggtaca aggcagatcc
cgatgagttt acctacgaga ggcgcaagga gtctgcagca tacattccct tcggcgaagg
ggattttat taccacgcag ccattttgg gggaacaccc actcaggtcc ttaacatcac
                                                                                       960
                                                                                      1020
                                                                                      1080
ccaggaatgc ttcaaaggaa tcctcaagga caagaaaaat gacatagaag cccaatggca
```

1140

tgatgagagc catctaaaca agtatttcct tctcaacaaa cccactaaaa tcttatcccc ggaatactgc tgggattatc atataggcct acctgcggat attaagcttg tcaagatgtc 1260 ttggcagaca aaagagtata atgtggttag aaataacgtc tga 1303

THIS PAGE BLANK (USPTO)

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the origina documents submitted by the applicant.

| efects in the images include but are not limited to the items checked | 1: |
|---|----|
| ☐ BLACK BORDERS | |
| ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES | |
| ☐ FADED TEXT OR DRAWING | |
| ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING | |
| ☐ SKEWED/SLANTED IMAGES | |
| ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS | |
| ☐ GRAY SCALE DOCUMENTS . | |
| ☐ LINES OR MARKS ON ORIGINAL DOCUMENT | |
| ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY | |
| | |

IMAGES ARE BEST AVAILABLE COPY.

☐ OTHER: _____

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)